DOCUMENT RESUME

ED 230 018 FL 013 133

AUTHOR Yorozuya, Ryuichi
TITLE Evaluation of Oral Fluency in English as a Foreign Language.
PUB DATE Jan 82
NOTE 121p.; Master's Thesis, Hiroshima University, Japan.
PUB TYPE Dissertations/Theses - Master Theses (042) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Audiolingual Skills; Communicative Competence (Languages); *English (Second Language); *Language Fluency; *Language Tests; *Speech Communication

ABSTRACT

Fluency can be defined in terms of intelligibility and effectiveness. For learners of English as a foreign language, fluency includes some degree of accuracy in phonological, syntactic, and semantic terms. It is a quality of speech that is produced in novel situations where what is said and how it is said cannot be predicted. To clarify the factors involved in the concept of fluency and to specify the kind of proficiency testing that would evaluate learners' fluency, a series of experiments were carried out. Ten adult non-native speakers of English were asked to describe the action in a comic strip. They were rated in two ways: (1) native speaker's subjective evaluation of the non-natives' fluency, and (2) non-native speakers' rate of speech and hesitation phenomena. A second experiment attempted to clarify the meaning of the term "non-native pauses." A comparison was made of descriptions of the same comic strip that were made by ten native speakers of English with the descriptions made by the non-native speakers. Non-native speakers were found to have more and longer hesitation pauses than native speakers. Also, the distribution was different. With regard to proficiency testing, the study points to the need for a workable alternative to ratings based on native speakers' intuition. (AMH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED230018

EVALUATION OF ORAL FLUENCY IN ENGLISH AS A FOREIGN LANGUAGE

Ryuichi YOROZUYA

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Ryuichi Yorozuya

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL013133

Hiroshima University Graduate School

EVALUATION OF ORAL FLUENCY IN ENGLISH AS A FOREIGN LANGUAGE

A Thesis
Presented to
Department of English Language Education
Specialized Curriculum Studies Major
Research Division in Education
Hiroshima University Graduate School

In Partial Fulfillment
of the Requirements for the Degree
Master of Education

---

by

Ryuichi Yorozuya

January 1982

TABLE OF CONTENTS

Preface ii

Chapter

O. Introduction 1

I. Communication and Speaking Ability 4

II. Fluency in English as a Foreign Language . . 16

III. A Survey of Current Techniques of Fluency Evaluation 41

IV. Experiment 47

V. Implications and Unsolved Problems 90

VI. Conclusion 98

Notes 101

Appendix 102

Bibliography 103

PREFACE

The desire to be able to speak a foreign language has attracted many to the study of English. However, it is frequently complained that it is difficult to improve one's speaking skills in a country like Japan. This is due to a number of unfavorable factors in the foreign-language learning context. These include the lack of suitable materials, teachers' lack of oral proficiency in English, teaching methods, class size, and so on. Perhaps one of the most serious hindrances to the development of oral proficiency, however, is the lack of appropriate techniques to test the ability to speak in the classroom. This problem of "how-to-test" stems from a lack of understanding of "what-to-test", i.e. "what it is to be able to speak a foreign language?". For every teacher and tester who wishes to assess speaking ability, it is fundamental to bear in mind what it means to be able to speak.

This study tackles this problem by focusing upon "fluency". Fluency is a characteristic and important feature of speaking ability. Without fluency, oral communication is seriously impeded. It is hoped that current tests of speaking ability might be improved and their goals clarified by the following heuristic attempt to offer a theoretical understanding of the nature of fluency and appropriate techniques for assessing it.

The present study is obviously limited in many ways and needs further refinement. Nevertheless, what is attempted here is intended to constitute the initial step in my research in Eigo Kyoikugaku "English Language Education," which will be for me a life-long pursuit.

My greatest debt of gratitude is owed to Professor Naomi Kakita at Hiroshima University, who introduced me to the field of Eigo Kyoikugaku and constantly inspired me to greater efforts, and Professor Mikio Matsmura at Hiroshima University, who has taught me the importance of deligence and caution in academic research. I must also thank Professor Kunio Okuda at Hiroshima University, who commented on the present thesis, and Mr. Toshiaki Ozasa at Hiroshima University, who offered me various suggestions about this thesis. Special thanks must go to Professor Toshinobu Kobayashi, who gave me an opportunity to use a pen-oscillograph for the data analysis in this study.

Thanks are due to those who were kind enough to serve as subjects for this study. The subjects included ten foreign students at the Intensive English Institute at the University of New Mexico, fifteen graduate and undergraduate students at the University of New Mexico, and ten Americans in Hiroshima.

I wish to thank the following friends and colleagues who gave me various suggestions: Mr. Tadashi Nishida, Mr. Yoshinori Nuibe, Mr. Tsuneo Nasu, Mr. Shigeki Kadoya, Mr. Tatsnori Takenaka, Mr. Seiji Fukazawa, Mr. Tomohiko Yanai,

Mr. Tōru Kinoshita, and Miss Keiko Koga. I also thank two American friends of mine, Mr. Derek Herforth and Mr. Donald Spence for reading the proofs of the present thesis.

Finally, I would like to express my deeply-felt thanks to Miss Chizuru Mori, Miss Patama Songpong, and Mr. Yoshikazu Yanagi for their help with typing and for their warm kindness and cheering encouragement. Any error found in the present thesis is my error and not theirs.

Ryuichi Yorozuya

January 1982.
Hiroshima.

INTRODUCTION

1. The Aim of This Study

The aim of the present thesis is to offer theoretical and empirical bases for effective evaluation of speaking skills. This thesis attempts to achieve this aim by clarifying the notion of fluency which is a sub-component of speaking skills and giving practical suggestions about how fluency should be evaluated.

2. The Scope of This Study

As the title of this study indicates, fluency is discussed with special reference both to "oral" (i.e. spoken) language and to learners of English as a foreign language. Thus fluency here does not refer to fluency in writing, reading, and listening or to the fluency of native speakers and speakers of English as a "second language" (as contrasted to a foreign language).

Another point that should be made is that the present study is primarily concerned with proficiency tests. Generally, language tests are classified into four kinds according to their functions (Davies 1968: 7): 1) Achievement test, 2) Proficiency test, 3) Aptitude test, 4) Diagnostic test. Achievement tests function as an integral part of a total educational system. They attempt to assess what and

how much has been learned of a known syllabus. Proficiency tests have the primary function of predicting a testee's future performance based upon a measurement of what he can do or knows now. They do not presuppose a particular syllabus, but do require the tester to have his own syllabus, that is, a clear idea of what is going to be tested. Aptitude tests also predict future performance but their primary function is to predict a testee's success in learning a language based upon assessment of his current first language skill. Diagnostic tests provide a teacher with information about what has gone wrong and how it should be treated in the future.

The reason for dealing only with the proficiency test in the present study is that if we were to consider other tests (e.g. achievement or diagnostic), we would have then to discuss fluency in relation to a particular course and this would cause an unnecessary complication in our subject matter. Rather, we will attempt to formulate our own idea as to what it means to be fluent in speaking English as a foreign language and consequently how fluency should be assessed.

The proficiency level of learners dealt with in the present study is intermediate and advanced. As Tsuchiya (1977: 12) points out, it is hardly possible to evaluate a learner's speaking ability unless he acquires basic skills to manipulate the phonological, syntactic and lexical elements

which enable him to convey meaning sufficiently. On this basis, fluency is judged to be a notion which applies to the intermediate and the advanced stages of language learning.

It should be understood that what is attempted in the present study is to deal with only one aspect of a complex pattern of linguistic behavior. However, any attempt to deal with one aspect of language behavior without considering other aspects and the whole would be invalid. Thus, whenever necessary, fluency is discussed in relation to the other aspects of communication, especially speaking.

3. The Organization of This Study

The present study consists of six chapters. Chapter I considers the process of communication and the act of speaking as the fundamentals of examining the nature of fluency. In Chapter II, basic consideration is given to the concept of fluency, its definition, and the identification of the specific factors involved. In Chapter III, current techniques for evaluating fluency are surveyed. In order to answer the three questions identified in Chapters II and III, a series of experiment was conducted. Those experiments are reported in Chapter IV. In Chapter V, implications for the evaluation of fluency are drawn from the findings of the experiment and some problems that remain unsolved are identified. Chapter VI contains the conclusion of this thesis.

CHAPTER I

COMMUNICATION AND SPEAKING ABILITY

1. The Importance of Speaking Ability in Foreign Language Teaching

1.1

Learning a foreign language puts one in possession of an important key to a door opening on a world one never knew of. It is the key to the mind of the peoples of the world. It is the key to open above all, the student's own mind. One way to obtain this key is to be able to "speak" a foreign language. The ability to speak has usually been one of the most important aims in modern foreign language teaching. There seem to be two reasons for this; one is from the student's point of view, the other from the teacher's point of view.

Anybody who learns a foreign language is bound to experience a strong desire to become capable of carrying on some conversation with a person who speaks the language. Rivers (1968: 161) points out that

> students come to the study of a foreign language in high school with the strong conviction that "language" means "something spoken".

Also we can find a number of empirical studies concerned with motivational aspects of students, and these show that students

are interested mostly in speaking a foreign language. Hatori and Matsuhata (1980: 199), for example, report the results of questionnaires answered by 2543 senior-high students: what interested most in learning English turned out to be "speaking". Similarly, Sawato (1980: 9)'s study reveals that a majority among 200 junior-high students are more eager to acquire speaking ability than the other three of the traditional four skills.

From the teacher's point of view, speaking is an essential aim in teaching English. Though there are some objections, the primacy of speech has been advocated on the grounds of the transferability of oral skills to the others. Lado (1961: 239), for example, mentions as follows:

> The ability to speak a foreign language is without doubt the most highly prized language skill, and rightly so, because he who can speak a language well can also understand it and can learn to read it with relative ease unless it happens to be a language like Chinese or Japanese, whose writing systems constitute major learning problems of their own. Also, the ability to speak a language will greatly expediate and facilitate learning to write it.

By putting an emphasis on speaking, teachers can expect some degree of positive transfer from speaking to other skills.

2. Models of Communication

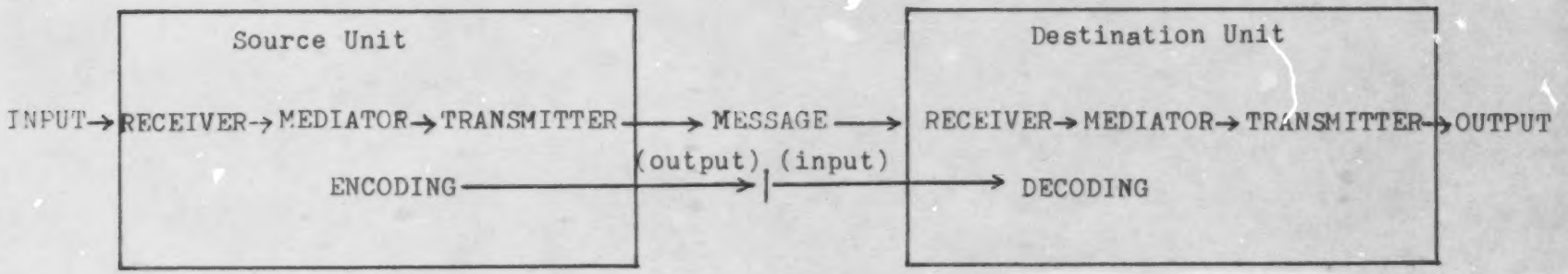
2.1

One's ultimate goal in speaking a foreign language is often expressed as "I want to be fluent in English." What does this term "fluent" mean? What is fluency in a foreign

language? This is not an easy question. Before we can attempt to answer this question we need first to answer the questions "What is communication?", "What is speaking?". The answers to these question are certainly premises for a clear understanding of the nature of fluency.

2.2

We will first examine three models of communication and then, in the light of the models, consider some aspects of the act of speaking. First, let us examine Osgood and Sebeok (1965)'s model of communication (Figure 1). Source Unit represents a speaker and Destination Unit a hearer. However, both units have the same apparatus which implies that the functions of each unit are interchangeable. Encoding process begins at Mediator where an intention is generated and the plan for its encoding is formulated. Then the plan is executed physically by Transmitter, an apparatus in the motor areas. This results into a message as output, which is in turn fed into the Destination Unit. First, the Receiver recodes the input into sensory neural impulses. Then the Mediator interprets them. From the model, thus, we can see how a communication process is initiated and terminated. In an everyday conversation, this process is chained together to form a long sequence of exchange, while in formal speech or in a lecture the process is completed through movement in only one direction. A monologue can be



(Adapted from Osgood & Sebeok 1965:3)

FIGURE 1

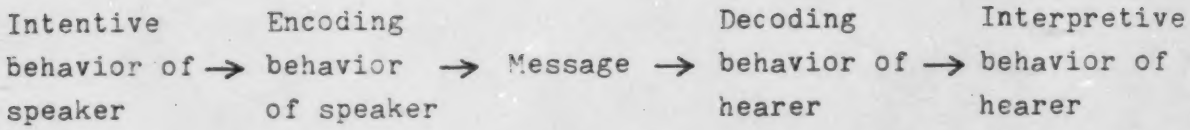
OSGOOD AND SEBEOK'S MODEL OF COMMUNICATION

regarded as a special case where the output of the Source Unit is immediately fed back into itself.

An advantage of their model is that it clearly demonstrates that the act of speaking is dependent on the other components, especially, the speaker's previous act of lestening. It can be seen from the process which takes place inside Source Unit that the act of encoding is closely connected with the Receiver and the Mediator. It should be remenbered that the ability to speak a language depends crucially upon the ability to comprehend. This is more clearly the case with the ability to speak a foreign language than with the ability to speak a native language. Later this point will be important in the discussion of the nature of fluency in English as a foreign language.

2.3

A similar model is presented by Carroll (1953: 88). The model consists of five stages (see Figure 2). The first two stages proceed on the speaker's side and the last two on the hearer's side. At the first stage the speaker conceives some kind of "information" which he wants to transmit. The linguistic encoding of the information is carried out at the next stage, viz, "encoding behavior of the speaker". This behavior is described as "a series of simultaneous and sequential 'choices' or 'decisions', made by the speaker, among one or more alternative learned responses in his repertoire, together with the



(Carroll 1953: 88)

FIGURE 2

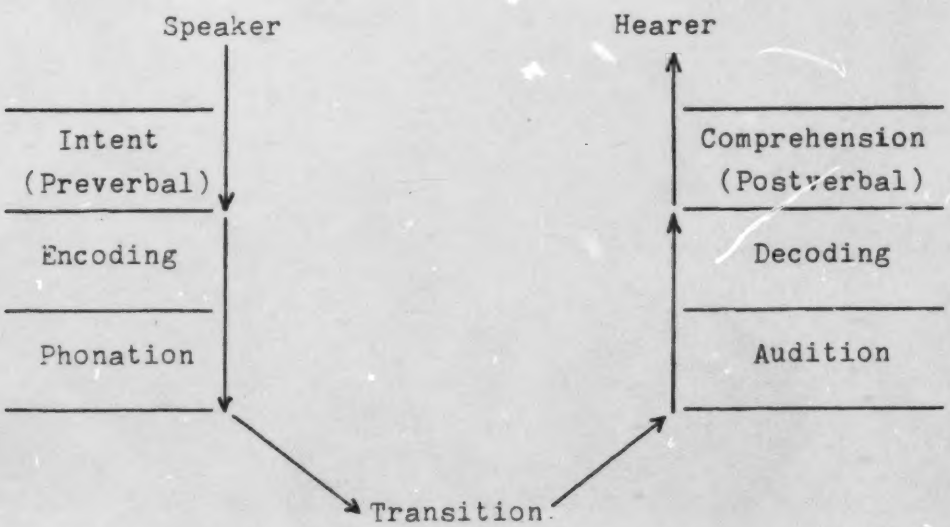
CARROLL'S MODEL OF COMMUNICATION

behavior of the motor systems (speech, gesture and so on) which each choice may entail"(Carroll 1953: 89). Thus, this stage involves both linguistic (phonology, syntax and lexis) encoding and physiological activity. Also the responses made here, Carroll points out, are "automatic but very complexly determined -- not only by the nature of the 'information' before it is coded, but also by associated responses, previous choice which have been made, the feedback from previous overt responses, situational factors, the nature of audience, and so on" (1953: 89). Thus, though this stage is only one of the five stages in the model and gives us a misleading impression of simplicity, it is easy to imagine how complex the decision makings involved in this stage are. Another point that should be mentioned concerning this stage is that the decision whether or not to make an overt response is also made here. Some speakers might have some information to convey, yet decide not to encode it. The output from the first two stages is a Message, which could be a succession of sound waves, written

marks on a page, or bodily movements. This message is then put into the next stage: decoding behavior of the hearer. Here, the message is perceived by the hearer and "a series of discriminatory responses" are made, that is, units of the language such as phonemes, morphemes and syntactical patterns are identified. Again the factors involved in this stage are multiple. The last stage, interpretive behavior of the hearer, is described as consisting of "information" which is not linguistically coded just like intentive behavior. At this point the hearer can turn into a speaker if the information evokes a response in him. Another sequence of the stages may thus be repeated.

2.4

The third model to be considered is Brooks'(1964: 4) model (see Figure 3).



(Brooks 1964: 4)

FIGURE 3

BROOKS' MODEL OF A LANGUAGE EVENT

His model is basically the same as Carroll's model. The main differences are that Brooks clearly divided Carroll's second stage (encoding behavior) into two: Encoding and Phonation, and the fourth stage (decoding behavior) into two: Audition and Decoding. Encoding and Decoding are concerned with linguistic processings while Phonation and Audition are physiological reactions.

2.5

Let us infer some fundamental features of communication from the models. These are that:

a) at least two participants are required: the sender and the receiver of message,

b) the two processes sending and receiving alternate and can be repeated,

c) some kind of information is transmitted from the speaker to the hearer,

d) both the speaker and the hearer must have a common linguistic code.

3. Basic Features of Speaking Ability

3.1

We can now turn to the consideration of the act of speaking seen as a part of the total communication process.

The act of speaking, basing upon Brooks' model, corresponds to the first three stages. First, some kind of information is conceived in the speaker's mind at Intent stage.

Then he decides whether to express it overtly. Sometimes this decision is made just before the linguistically coded information is actualized by the physiological movements of the articulatory organs. Referring to Brooks' model of communication, the decision can be made either before Encoding or Phonation stages. Encoding stage involves giving the information a linguistic form: suitable words and structures are assigned. The resultant linguistic plan is next actualized by the Phonation process.

3.2

There are five points to be made about this act of speaking. First, it should be emphasized that numerous factors are operative at each of the stages Intent, Encoding and Phonation. The shape of an end-product which has gone through these three stages and the way it is articulated should be seen as determined multiply by those factors. Those factors include the previous speaker's response, the situation in which the act of speaking takes place, the speaker's knowledge, register, interpersonal relationship with the hearer, the culture which dominates to some extent what is said and how it is expressed, the nature of the audience, verbal or non-verbal feedback by the audience, and so on. Every element involved in the total communication contributes potentially to the act of speaking.

Second, the processing which takes place is not linear

but hierarchical in nature as Rivers (1968: 162) points out. The sequence of the sounds generated from the speaker's mouth gives us an impression of being linear since the sounds and words come out one by one. However, in the speaker's mind the process of generating a sentence is by no means linear. A decision at one level demands the matched decision at the next level and this decision further determines the decisions at other levels. A particular intention of a speaker entails limited overall structures. The intention and the structure affect the choice of the words in the sentence. The words determine which sounds should be arranged and articulated. A number of decisions are made and carried out at several levels simultaneously when one speaks.

Third, most of these decision-making processes and motor activities have to be done without much awareness. If the speaker must be conscious of all the elements involved in speaking, he cannot hope to speak efficiently. (The classic example is the non-native foreign language teacher who is too meticulous about the accuracy of his grammar and pronunciation.)

Fourth, the act of speaking should not be regarded as separable completely from his act of listening and the listener's reaction to his speech. As was mentioned in connection with the first point made above, prior to the speaker's speech-act, his reaction to what was said to him is

formed, unless, of course, he initiated the conversation or he is lecturing or making a speech. His reaction to the previous speech must be taken into account when he decides what to say next and how to encode it linguistically. Osgood and Sebeok (1965: 3)'s model described above clearly shows this point, incorporating Receiver in Source Unit. Also one's act of speaking cannot be separated from one's listener in the sense that the speaker constantly checks how the listener reacts to his speech. Stevick (1976: 63) quotes an interesting experiment by Blubaugh (1969) in which one group of speakers were exposed to an audience who reacted to their speech in quite a favorable and positive way nodding and smiling while the others were exposed to those who reacted in a negative way avoiding eye contact and fidgeting. The latter group performed poorly in making their speeches.

The last point to be made concerning the act of speaking is that the speaker cannot escape from time constraints. He is not allowed in conversation to think about what to say and how to say it for one hour and then begin to speak. Crystal and Davy (1975: 87) state that the main factor which distinguishes written from spoken language is time. Indeed this is one of the most salient features of the act of speaking and is closely related to the concept of fluency.

Having considered the processes in communication and the various features of the act of speaking, we now turn to

the elucidation of the nature of fluency.

CHAPTER II

FLUENCY IN ENGLISH AS A FOREIGN LANGUAGE

So far we have considered what happens in a communicative situation and what is involved in the act of speaking. Our next question is "what is it to be fluent in English as a foreign language?"

1. The Problem

1.1

"Fluency" has been and will no doubt continue to be an abused word. People use the word differently in various contexts. It has been widely used to express the ultimate goal of modern language courses, or as a criterion for testing the ability to speak. Madsen and Jones (1980: 7), for example, report on a survey of seventy-four speaking tests, both published and unpublished. Their finding was that of the seventy-four tests the most frequently used criterion was Grammar (81%), the second was Fluency (71%), and the others were Vocabulary (67%), Pronunciation (66%), Appropriateness (63%), and Others (37%). Though it is conceivable that the definition of fluency might vary from one test to another, we can see how widely the term fluency is used as a feature used measuring the ability

to speak.

1.2

However, the definitions of fluency are frequently ambiguous. Ingram (1968: 78) states as follows:

> There has been a certain amount of not very fruitful discussion in recent years about 'fluency' both from the teaching and the testing points of view. This is not surprising, for 'fluency' is not a defined term, it is just a label, giving a misleading impression of simplicity to that very complex thing, command of spoken language.

We could refer to a dictionary for the general consensus on the definition of fluency. Webster's Third New International Dictionary of the English Language (1966) defines 'fluency' as being "versed in the use of language: ready with words . . . easy and flowing: pleasingly graceful." The Concise Oxford Dictionary defines it as the quality of being "copious, coming easily, ready: expressing oneself quickly and easily." It seems that the general agreement in these definitions is that fluency is a quality of being able to produce a considerable quantity of language within a certain time with ease and smoothness. However, these definitions are too vague to be used as bases of testing fluency, as exemplified by such words as "easily" and "readily". In order for us to establish a valid technique of measuring fluency, it is essential to have a more precise and concrete definition of this key term.

2. Some Preliminaries

2.1

It is necessary here to have a preliminary discussion about some points concerning the concept of fluency. Three points will be discussed: the problem of meaning in fluency, production and reproduction, and the difference between native and non-native speakers' fluency.

2.2

Can a person be called fluent if he cannot make the listener understand what he means? Many would answer this question in the negative. From the stand-point of a third person unacquainted or minimally acquainted with the language being spoken, the speaker may look as if he were conveying some meaning, although, in fact, he is merely performing a vain, meaningless charade. This is "pseudo-fluency". In this study, we will not regard this kind of speech as true fluency. It must be emphasized that the transmission of meaning is fundamental to the concept of fluency.

In order to elaborate this point, let us refer to Catford (1950)'s distinction between <u>intelligibility</u> and <u>effectiveness</u>. According to Catford, the two terms are defined as follows:

> Speech is generally said to be <u>intelligible</u> if the hearer 'understands the words', <u>i.e.</u> if his response is appropriate to the <u>linguistic forms</u> of the utterance: that is to say, if it is in accordance with the semantic habits of the speech-community whose language is being used. An utterance may be intelligible in this sense, yet <u>ineffective</u> in the

> sense that the hearer's response is not what the speaker intended. (Catford 1950: 8)

Following this, he shows the four possible combinations of the two.

TYPE OF UTTERANCE	HEARER'S RESPONSE
(1) Unintelligible+Ineffective	None, or inappropriate
(2) Intelligible+Ineffective	Inappropriate
(3) Unintelligible+Effective	(Appropriate, only by chance)
(4) Intelligibe+Effective	Appropriate

(Catford 1950: 8)

Type 3 is disregarded completely since it does not depend on an effective use of a foreign language but haphazard means of communication such as gestures, facial expressions, pictures, and so on. Type (1) and (2) are described as unintelligible, and (4) as intelligible. Type (4) is considered as important since "the language learner's speech should be not only intelligible, in the narrow sense, but also effective" (Catford 1950: 9).

To be qualified as fluent, type (4) is the most desirable, and type (1) and (3) are unacceptable. Type (2) is a marginal case since in type (2) the ineffectiveness might not totally be due to the speaker. An utterance could be judged ineffective because of a fault on the part of the hearer. Here, Catford (1950: 10)'s five levels of loss of intelligibility are worth quoting:

> (1) Speaker's <u>selection</u> of linguistic forms.
> (2) Speaker's <u>execution</u> of linguistic forms.
> (3) <u>Transmission</u> from speaker to hearer.
> (4) Hearer's <u>identification</u> of linguistic forms.
> (5) Hearer's <u>interpretation</u> of linguistic forms.
>
> (Catford 1950: 10)

Though he originally set up these levels in order to show at which points of communication loss of intelligibility may occur, we could also apply them to loss of effectiveness. Presumably, the speaker is responsible for failures at points (1) and (2): at point (1), he fails to choose an appropriate word, i.e. one that accurately expresses his intention: at point (2), he fails to execute the 'mechanics' of speech as intended (e.g. mispronunciation, false-starts and so on). The points (3), (4), and (5) do not lie in the speaker's domain. Thus, we do not regard a speaker as non-fluent even if some failure occurs at points (3), (4) or (5). This is because the concept of fluency is defined primarily in terms of the speaker. In short, if one wants to be qualified as a fluent speaker, first his speech has to be intelligible, second the effectiveness of his speech should not be lost at either of the first two of Catford's five levels.

Concerning intelligibility, some qualification is necessary. We stated that fluency presupposes intelligibility. Intelligibility in turn presupposes some degree of accuracy in phonological, syntactic and semantic terms. Consequently, it would follow that the concept of fluency includes "accuracy". However, we should bear in mind for later discussions that this only means that fluency presupposes some degree of accuracy enabling a learner of English as a foreign language to make him- or herself understood by native speakers of English, that

is, to be at least intelligible. It is obvious that we cannot determine precisely what degree of accuracy is really needed in terms of phonology, syntax and lexis in order to be at least intelligible since there is always some help from other elements in a communicative situation and the relative importance of each element is subject to change. However, it is not too difficult to judge the intelligibility of one's speech as an overall effect of smaller elements of his speech. The term "accuracy" which is used in later chapters therefore refers to higher levels of accuracy than the level of accuracy at which speech is minimally intelligible.

The reason for regarding unintelligible speech as not fluent is that, as Tsuchiya (1977: 12) maintains, it is difficult to measure the ability to speak unless the speaker has acquired some degree of accuracy in using English as a foreign language. This is also why we excluded elementary learners from our consideration, as stated in the Introduction.

2.3

Katz and Fodor (1964: 481-482) as follows:

> What qualifies one as a fluent speaker is not the ability to imitate previously heard sentences but rather the ability to produce and understand sentences never before encountered. The striking fact about the use of language is the absence of repetition--almost every sentence uttered is uttered for the first time.[1]

It should be emphasized that in the present study fluency in English as a foreign language is not used to describe one's

speech as a reproduction of previously given text.

Matsuhata (1974: 8) distinguishes two kinds of act of speaking: production and reproduction. He maintains that the act of speaking can be characterized as a social activity in which one cannot predict what will be said by the other person. It is emphasized that to be able to answer a question "How old are you?" in the classroom and to be able to respond appropriately to questions asked in a real communicative situation are quite different. Indeed, no matter how native-like one's speech might be in a situation where he, for example, recites a story which has been rehearsed a number of times, we cannot tell whether he is really fluent.[2] It is only when he is required to speak in a novel situation that we can truly judge whether he is fluent or not. Thus, in order to evaluate one's fluency, testing situations and speaking tasks should be chosen carefully. Fluency is a quality of speech which is not rehearsed but rather produced in a novel situation where what should be said and how it should be said cannot be wholly predicted.

2.4

The distinction should be maintained between fluency of a native speaker of English and that of a non-native speaker. The distinction could either be qualitative or quantitative.

Native speakers are assumed to have a certain basic language competence. Their command of the native language is

supposed to surpass a common minimum level. Fluency used to describe a person's native language use therefore refers to a quality of speech which is above that minimum level commonly achieved by all native speakers. In this sense fluency in the native context could be regarded as belonging to performance rather than competence. To exemplify this a native speaker is commonly said to be fluent if he uses rhetoric skillfully and sounds "pleasingly graceful" (Webster's Third New International Dictionary of the English Language: 1966), or if what he says is well-organized. Though these performance factors do influence our judgement about a non-native speaker's fluency, these factors are subordiate to other major factors which we will discuss later.

Another difference between native and non-native fluency stems from the above argument. The degree of efficiency in language use which a non-native speaker has to attain in order to be qualified as a fluent speaker is distinctly lower than that of a native speaker. Consider Words per Minute, for example. Carroll (1964: 60) says that the rate of normal uninterrupted English speech is about 165 words per minute or 265 syllables per minute. In the case of Radio broadcasting the average rate of speech is about 190 words per minute for domestic broadcasing in the United States and about 180 for that in Britain (Kaise 1980: 17). In order for a native speaker of English to be qualified as fluent, he probably has

to speak at a speed faster than about 165 words per minute, to use Carroll's figure. However, a non-native does not have to speak at such a fast speed to be qualified as a fluent speaker of English as a foreign language. Presumably the rate of speech required for a non-native speaker is approximate to 165 words per minute or lower.

In short, fluency in the native context is used to describe a degree of linguistic competence distinctly superior to the normal level of efficiency, while in the non-native context, the same word points to a level of competence close to the native speakers' norm. The two kinds of fluency naturally overlap to some extent, but they also differ in many respects.

3. The Concept of Fluency in Foreign Language Teaching

3.1

Let us first define the concept of fluency so that the definition can be used in devising an appropriate technique to measure fluency. In the following discussion one of the main issues is the relationship between fluency and accuracy.

In defining fluency, it will be useful to consider Dalton and Hardcastle (1977: 2-3)'s and Leeson (1975)'s opinions. Dalton and Hardcastle (1977: 2-3)'s work is primarily concerned with fluency disorders in native speech.

However, their analysis of the concept of fluency generally held in the field of foreign language teaching is worthy of consideration. They state:

> There seems to be more involved in an assessment of fluency in a foreign language than the temporal and sequential aspects of the speech process. Accuracy in the use of the language is another potentially important factor to consider and will include such features as correct adherence to the phonological, syntactic and semantic rules of the language, to prosodic variables such as stress and intonation patterns, as well as to the use of appropriate vocabulary. Thus the main factors contributing to fluency in the foreign-language-learning context can be grouped under two main headings: those relating to accuracy in the use of the language, and (somewhat more specifically) those relating to temporal and sequential aspects of the speech output, such as repetitions and pauses. (Dalton and Hardcastle 1977: 2-3)

Allowing for apparent oversimplification, we can represent their idea as follows:

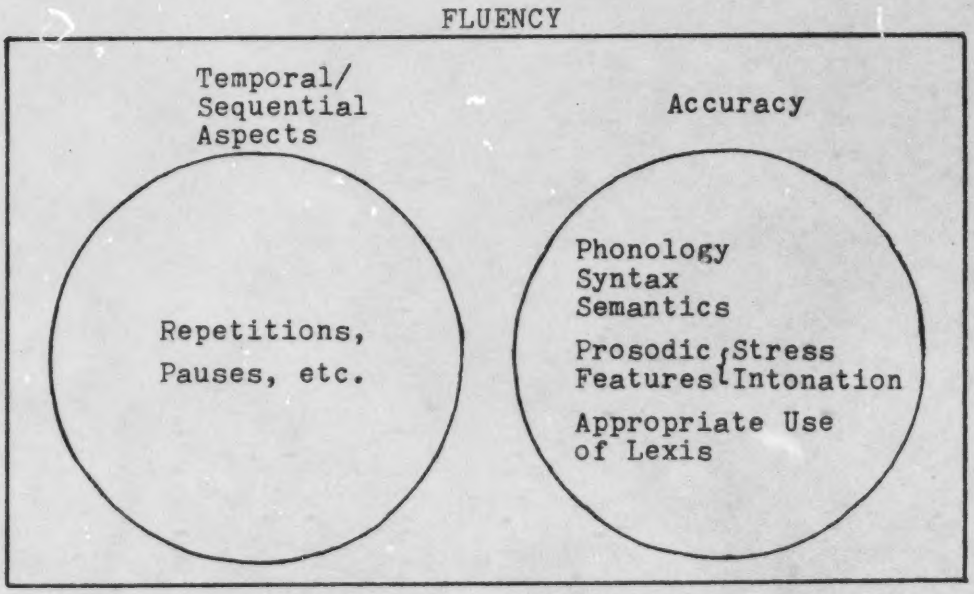


FIGURE 4

GENERAL CONCEPT OF FLUENCY IN THE FOREIGN-LANGUAGE-LEARNING CONTEXT BY DALTON AND HARDCASTLE(1977)

The outer square embraces factors involved in the concept of fluency. The blank space within the square but outside of the circles does not include any factor if we adhere strictly to their statement, but it is possible for us to include in this space some factors which belong to neither of the circles, such as the organization of what is being said and paralinguistic phenomena such as gestures, facial expressions, and so on.

As Dalton and Hardcastle point out, "fluency is used to describe a level of efficiency in a foreign language," incorporating temporal / sequential aspects of speech output and accuracy. Thus fluency in their definition covers a wide spectrum of linguistic skills.

A similar idea about fluency has been suggested by Leeson (1975). According to him, fluency is defined as "the ability of the speaker to produce indefinitely many sentences conforming to the phonological, syntactical and semantic exgencies of a given natural language on the basis of a finite exposure to a finite corpus of that language (Leeson 1975: 136). In this definition "productivity" is emphasized, and this, of course, implies that he has been influenced by the current theoretical trends in linguistics. He states that the definition above is "the fundamental criterion against which may be measured the performance of the individual speaker. . ." What he means by fluency is therefore almost equivalent to

the ability to speak, to language competence which involves the capacity to learn a language. Based on this definition, Leeson offers some tentative criteria of fluency (pp 143-146). These include three components: phonological, syntactic, and semantic. For each of the three, several criteria are listed:

1. Phonological Component
   a) Articulation rate
   b) Pausal phenomenon
   c) Phonological production
   d) Phonological discrimination

2. Syntactic Component
   a) Error count
   b) Gap filling
   c) Syntactic manipulative skill
   d) Anagram sentences

3. Semantic Component
   a) Associative networks in a variety of registers, socio-cultural influences and referential elements common to the society of the individual.
   b) Awareness of the functional value of utterances.

If compared with Dalton and Hardcastle (1977)'s idea, it can be seen that their temporal / sequential aspect is incorporated in Leeson's phonological component and accuracy in all three of his components. A slight difference is that Leeson incorporates phonological discrimination in his scheme, based on his recognition that fluency partly depends on the speaker's comprehension. Throughout his book it is emphasized that fluency cannot be assessed without taking into account all those aspects described above.

Our definition of fluency in the present study is narrower than both Dalton and Hardcastle (1977)'s and Leeson

(1975)'s. Referring to Dalton and Hardcastle's two categories of factors in fluency, we regard fluency as relating primarily to the temporal and sequential aspects of speech output. This definition is based upon the observation of current definitions of fluency used in the testing speaking skills. As will be described in Chapter III, most of these definitions are used to standardize native judges' evaluation of fluency in interviews, directed conversations or monologues (Clark and Swinton 1979: 66, Harris 1969: 84, Valette 1977: 159, Clark 1972: 93). We can regard them as representing current thinking in the field of foreign language testing. In all of these definitions, prime attention is paid to temporal and sequential aspects of speech output. Since these definitions are for actual use in testing situations, it can be said they represent the most practical kind of definition. Our initial assumption in the present study is that fluency is one component of the speaking skills. As the present study attempts to devise an appropriate technique to evaluate that one component, Dalton and Hardcastle's and Leeson's definition of fluency are judged to be too broad. Our aim here is to try to compartmentalize the speaking skills and offer some specific criteria for a valid and reliable measurement of fluency.[3]

However, despite our definition of fluency as a relatively narrow concept which relates primarily to temporal and sequential aspects, it is undeniable that fluency in this sense

is nevertheless related to some aspects of accuracy, especially when native speakers try to evaluate a learner's fluency according to our definition. It would seem that phonological and prosodic accuracy in particular influence native speakers' evaluation of fluency in the sense defined above. Thus, though we do admit the close relation of fluency in our sense to various factors in accuracy of speech output and that it might be difficult to distinguish the two components in practical terms, we still maintain the dictinction between the two components at the theoretical level.

3.2

What factors, then, are involved in fluency as defined above? We classify them into two categories: 1) those factors which are observable in the act of speaking and are overt evidence of fluency and 2) those factors which affect the learner's fluency internally, that is, factors which work together to determine his level of fluency.

Let us consider the factors in the first category. When one speaks a foreign language, the level of fluency can actually be judged from the following factors:

1) Speed of response
2) Length and number of hesitation pauses
3) Number of repetitions and false starts
4) Logical coherence and the use of connectives

These factors thus represent in concrete terms what is meant by the "temporal and sequential aspects of speech output", which characterize the nature of fluency as defined in the present study. It is proposed that the evaluation of fluency in English as a foreign language should be based primarily upon these four factors, though they may not be exhaustive.

"Speed of response" includes two components: the duration of the "latent" period after one speaker finishes and before the next speaker begins to speak, and, secondly, the rate of speech of the speaker being evaluated. These two are considered to be the most crucial in the evaluation of fluency. Both are closely related to the learner's efficiency in encoding. The first is also dependent upon his efficiency in decoding what was just said.

It should be remembered that no matter how rapidly one makes a response he cannot be judged to be fluent if his response does not make sense in the context. Stack (1966: 186) regards this kind of speech as poor, giving the following example:

S: What's your name?
Student: Fine, thank you. (<u>immediate answer</u>)

This point was emphasized earlier in this chapter.

Concerning the second component, the rate of speech, some qualification is necessary. Common indices used for this aspect of fluency are "Words per Minute (WPM)" or

"Syllables per Second (SPS)". Though these are perhaps the most objective measures of fluency, they are not wholly satisfying as fluency indicators. Since mere numbers of words or syllables are used as the bases for evaluation, these indicators fail to take into account the varying amount of information contained in a particular word or syllable. Indeed, what matters in the measurement of fluency should not be simply the superficial amount of language produced in speech output. If one actually hopes to measure a foreign language learner's fluency with the greatest possible accuracy, he should take into account the total amount of information which is in accord with the given context. Suppose that A and B are at a hamberger shop and both of them want to take the hamberger back home. When asked whether they want the hamberger to take out or to eat on the premises, A answers "Well, I want to take it home and eat it there", while B answers "To go, please". Suppose further that both of them spoke at 150 WPM. If we used WPM as the sole criterion of fluency, we will naturally judge that A and B have exactly the same level of fluency. But is this accurate? From the standard of amount of information conveyed within a certain time, B is naturally more fluent.

This does not, of course, mean that the more information one can convey through speech, the more fluent he is. As natural languages require a certain degree of redundancy, especially in the spoken mode, there certainly are upper and

lower limits on the amount of information conveyable within a certain time (Rivers 1968: 137-8).

We could use a measure of the amount of information conveyed within a certain time as an index of fluency. The technique developed by Goldman Eisler (1968: 34) based on Shannon (1951) could be used. However, this technique for estimating the amount of information conveyed is quite involved and impractical as a means of measuring fluency. Thus, admitting that WPM is only a crude indicator of fluency, we nevertheless adopt it here as a convenient way allowing us to measure relative fluency with a fair degree of accuracy.

Length and amount of hesitation pauses can "constitute one of the bases for a meaningful evaluation of fluency" (Leeson 1975: 76). It is intuitively clear that hesitation pauses affect the degree of fluency as judged by native speakers. The longer and more frequently one pauses the less fluent he sounds.

The third factor would also appear to be a significant criterion for measuring fluency. Repetitions in particular, are apt to interrupt the smooth flow of speech when a foreign language learner speaks. These are frequently accompanied by long pauses, indicating that repetitions occur as a devise to allow some time for finding the next word or to correct the previously uttered word.

The forth factor, logical coherence and the use of con-

nectives is probably the most difficult to measure accurately. It has been suggested by Crystal and Davy (1975: 86-107) that fluency (the continuity of utternace) in an informal conversation is maintained by the use of connectives. They list four kinds:

1. neutral connectives "and....and....and"
2. reinforcing or supplementing connectives
   "in fact, in other words, for instance, etc."
3. diminishing connectives
   "at least, or rather, etc."
4. softeners "you know, I mean, sort of, you see."

Appropriate connectives in a suitable context would presumably make one's speech smoother and easier to follow. As the communicative effects of such connectives is not precisely understood, there are practical difficulties with using connectives as indices of fluency. Further research is needed especially on the communicative effect of particular types of connectives such as the four proposed by Crystal and Davy (1975).

The correct use of connectives implies that the learner's speech is logically organized. This again cannot easily be quantified. Though this factor is merely listed here as a possible candidate for an index of fluency and no attempt to implement this consideration in practical terms is made in the present study, it is nevertheless recognized that if a complete measurement of fluency is aimed at the logical coherence of speech output cannot be neglected completely.

The four factors listed and described above are proposed

as constituting a significant part of the overt evidence from which native speakers judge a foreign language learner's level of fluency. Thus, they can be useful criteria for measuring fluency.

In Chapter IV, we will be concerned with the verification by experiment of the validity of some of these factors as criteria for measuring fluency. We will examine particularly the validity of using 1) Speed of response and 2) Length and amount of hesitation pauses as criteria for measuring fluency, and their relative significance will be evaluated.

3.3

Next, we consider the factors which internally affect the level of the learner's fluency as the temporal and sequential aspects of speech output. Put differently, the question we try to answer here is "What factors internally determine the level of fluency?" As was briefly discussed in Chapter I the act of speaking is influenced by numerous factors inherent in the speaker, the listener(s) and the whole communicative context. It is the same with fluency. This is obvious since fluency is an aspect of the act of speaking.

Though far from exhaustive, the following factors seem to influence the level of fluency:

1. Speaker factors

(a) Encoding capacity (phonological, syntactic, semantic)

(b) The efficiency of articulatory organs

(c) Knowledge and experience

(d) Motivation to speak

(e) Personality

(f) The psychological state

(g) Comprehension

2. Topical factor

3. Cultural factor

4. Situational factor

5. Listener factor

The first factor pertaining to the speaker - (a) Encoding capacity - is most crucial in determining the learner's level of fluency. Actually, this factor has two aspects: the efficiency of language processing and the size of the repertoire which the speaker possesses. These two aspects can respectively be devided into phonological, syntactical, and semantic components. The first aspect can be said to belong to the domain of performance. The efficiency of language processing determines how fast one can encode a given intention. To refer to the competence-performance dichotomy, encoding requires not only competence, i.e. knowledge, but also performance, i.e. facility. What is meant by the first is for example, how fast one can build up a sentence which matches a certain idea, or remember a specific word referring to an object or an idea. This is a cognitive capacity and thus should be distinguished

from Speaker factor (b), the efficiency of articulatory organs. The second aspect of this factor, the size of the linguistic repertoire possessed by the speaker, can also be looked at in terms of phonology, syntax, and semantics, but this aspect of the encoding capacity is most closely related to the semantic domain, especially the number of vocabulary items the speaker has at his disposal. In order to be fluent, the speaker has to have a fairly good so-called "active vocabulary". The importance of semantic component in the ability to speak has been emphasized by Leeson (1975: 102), who states that "the semantic component is, perhaps, in the final analysis, the key factor in differentiating between the performance of one individual and another". This will be keenly felt when we analyse hesitation pauses, one of the crucial factors in fluency. We notice that a cause of hesitation pause is the speaker's difficulty in choosing vocabulary or his lack of vocabulary.

One point that has to be emphasized here is that, as can be seen from the above argument, fluency is actually a composite made up of factors concerning phonology, syntax, semantics. This implies that the ability to operate these three components might be closely connected with fluency.

The efficiency of articulatory organs also influences the level of fluency. No matter how rapidly the cognitive processing is done and linguistic plans are formulated, if the

speaker's articulatory organs are not efficient, the resulting level of speech output is low. This is especially the case with foreign language learners who have not been accustomed to pronouncing the sounds of the language they are learning.

A foreign language learner's level of fluency depends largely on his knowledge and experience, though these are something which language testers would like to eliminate from the measurement of fluency. Even the most proficient learner could be judged unfluent in talking about, say, politics, if he knew nothing about politics. A closely related factor is one's familiarity with the particular expressions unique to a register. This point has been pointed out by Leeson (1975: 127), who states that

> in those registers where the speaker was highly experienced, the syntactic structures upon which part of the fluency of ideational expression depends would feature in his semantic field for that register as more or less conditioned structure groups with multiple alternative lexical options ready to be run off as required.

One's true fluency in a foreign language is thus sometimes difficult to be clearly distinguished from his knowledge and experience, and his familiarity with a particular register. This suggests that testing tasks for measuring the fluency of foreign language learners should be those which do not require specialized knowledge or experience. What we are primarily interested in is learners' fluency as their linguistic efficiency. Therefore, testing tasks should be chosen carefully.

Motivation to speak strongly affects the intentive behavior of the speaker in Carroll (1953: 88)'s model of communication, thus consequently the speaker's fluency.

Personality factor such as extroversion / introversion are also supposed to be related to fluency. However, as accurate measuring tools are still in the course of development (Naiman, et. al. 1978: 67), the precise relationship between fluency and personality is not yet clear.

The factor (f) most acutely shows itself in a live speaking test such as an interview. This is indicative of the importance of setting up a relaxed testing situation when fluency is being measured.

The last speaker factor is the speaker's capacity to comprehend what is said to him. This factor is most apparently operative in the conversational situation. Of the two aspects of the first factor, Speed of Response, stated in the previous section, the amount of latency before the speaker begins his speech depends largely on this factor. Thus, fluency should be recognized as concerned not solely with the speaker's encoding capacity but also with his decoding capacity.

The following three major factors are separated from Speaker factors since they do not exclusively affect the speaker.

The topical factor obviously has much to do with Speaker factor (c). However, regardless of the speaker's knowledge and

experience, there presumably are topics universally considered difficult to talk about. The fact that fluency is affected by this factor has clearly been demonstrated by Goldman Eisler (1968).

The cultural backgrounds of learners is also relevant to their level of fluency. Those learners whose background culture is the one in which "silence is gold" might be less fluent than others, since their native habit to be humble would inhibit them from saying very much.

The foreign language learner, though less sensitive than native speakers, is presumably aware of situational requirements. Whether he is talking in an informal conversation with his friends or in a formal speech at a dinner party would affect the rate of speech, the organization of what is said, and so on.

The last factor, Listener factor, is related to fluency in two ways. First, the listener's social rank and age are likely to influence the speaker's fluency. If the speaker is talking to a man of importance, he is supposed to use a more elaborate speech level and, as a result, his fluency will be affected. Secondly, as was mentioned in Chapter I, the listener's feedback through various means sometimes influences the speaker's fluency.

All these factors directly or indirectly determine the level of fluency. What is important here is that, though we

have proposed that fluency can be measured by considering four major external features of speech output, fluency is an aspect of speaking ability which is affected continuously by numerous factors existent in the whole communicative situation. This integrative nature of fluency has been pointed out by Oller.[5] In the present study, this will be verified in the experimental section by investigating the degree of the integrative nature of fluency as compared with others of the speaker's linguistic capacities - pronunciation, grammar, and vocabulary.

CHAPTER III

A SURVEY OF CURRENT TECHNIQUES OF FLUENCY EVALUATION

In this chapter we will survey some of the testing techniques for evaluating fluency in speaking a foreign language and identify some implications of current testing techniques. This will lead to some modest suggestions for improvement of fluency evaluation methods.

1. Tests of Fluency

As the concept of fluency itself is not clearly defined and a general consensus about it is lacking, each tester constructs a testing battery from his own interpretation of this basic concept. Testing techniques seem to be divisible into two types: those which take into account the quantity and accuracy of language produced and those which take into account the temporal and sequential aspects of speech output. The first of these adopts the interpretation of fluency proposed by Dalton and Hardcastle (1977), that is, the idea that fluency consists of temporal and sequential factors and accuracy factors. The second considers fluency as relating to temporal and sequential aspects of speech output. This latter view

has also been followed in the present study.

An example from the first category is provided by Carroll (1979). His technique of measuring fluency was developed as a part of the International Association for the Evaluation of Educational Achievement (IEA)'s International Study of Achievement in French as a Foreign Language. In the section "Speaking Tests", there were separate Pronunciation, Structural Control, Oral Reading, and Fluency Tests. Carroll states that

> the Fluency test was designed to obtain information about the quantity and quality of the student's response in French when he was asked to speak somewhat more at length in describing sequences of events or in pictured situations. (Carroll 1979: 16)

The scores produced by him were as follows:

> a=Number of intelligible clauses ("propositions," in French; either an independent or a dependent clause; scores [b] through [f] were based only on these "intelligible" clauses);
> b=Number of different grammatical structures represented in the response;
> c=Number of clauses correct in structure;
> d=Number of clauses correct in morphology (inflection);
> e=Number of clauses correct in vocabulary;
> f=Number of clauses correct in pronunciation.

Then the relative importance of each score was determined by using global ratings of 1933 examinees' responses, and the resulting scoring formulas were as follows:

> Fluency Quantity Score=2b+2c+d+2e+2f;
> Fluency Quality Score=(6c+3d+6e+6f)/a;
> Total Fluency=Fluency Quantity Score+Fluency Quality Score, where a,b,c,d,e, and f are as defined above, and scores were set to zero when variable a=0

There is an interesting point about this test. Only

intelligible clauses are considered in determining the level of fluency. This point is in complete accord with the statement we made earlier that meaning is fundamental to the concept of fluency. This point seems to be the most essential in any evaluation of fluency.

As can be seen from the scoring formulas, Carroll takes into account of correctness (accuracy) and the amount of speech output. In this regard, his test of fluency samples the factors involved in the ability to speak fairly comprehensively. In addition, the scoring procedure is objective, using quantifiable measures, so that the reliability of the test is enhanced. Thus we can say that his test is well-designed. However, fluency as defined in the present study - temporal and sequential aspects of speech output - is not separately measured in the test. What Carroll means by "fluency" is equivalent to "the ability to speak." Though his idea about fluency overlaps partially with ours as defined earlier, we regard his test of fluency as a test of the total speaking ability rather than as a test of fluency.

Next, we will examine tests which assess fluency as only a part of the total speaking ability. There are a substantial number of tests of this sort. A common procedure is to have native speakers of the language being tested evaluate a foreign language learner's speech in an interview, directed conversation, or monologue using as an evaluative

scale a set of definitions describing four or five levels of fluency. This type of test of fluency has so far been the most prevalent. Some examples of definition of fluency levels are given below.

1) The Foreign Service Institute Oral Interview, Numerical Rating Procedure: Fluency

1. Speech is so halting and fragmentary that conversation is virtually impossible.
2. Speech is very slow and uneven except for short or routine sentences.
3. Speech is frequently hesitant and jerky; sentences may be left uncompleted.
4. Speech is occasionally hesitant, with some unevenness caused by rephrasing and groping for words.
5. Speech is effortless and smooth, but perceptibly non-native in speed and evenness.
6. Speech on all professional and general topics as effortless and smooth as a native speaker's.

(Valette 1977: 159)

2) Foreign Service Institute "Descriptive Statements" Rating Criteria: Fluency

1. Speech is as smooth and as effortless as that of a native speaker.
2. Speech shows only occasional evidence of "non-native" pausing.
3. Numerous non-native pauses in course of interview.
4. Speech is so halting and fragmentary that conversation is virtually impossible.

(Clark and Swinton 1979: 66)

3) Harris' Oral English Rating Sheet: Fluency

5. Speech as fluent and effortless as that of a native speaker.
4. Speed of speech seems to be slightly affected by language problems.
3. Speed and fluency are rather strongly affected by language problems.
2. Usually hesitant; often forced into silence by language limitations.
1. Speech is so halting and fragmentary as to make conversation virtually impossible.

(Harris 1969: 84)

1) The FSI's numerical rating procedure is used to supplement global ratings using five levels of proficiency which are assigned in an interview. The numerical ratings are assigned to Accent, Grammar, Vocabulary, Fluency, and Comprehension.

2) The "Descriptive Statements" are not part of the regular FSI system but were developed specially by Clark and Swinton (1979: 38) for their project "with the goal of expressing, in somewhat more operational terms, the same general kinds of discrimination sought" in the regular descriptive statements used in interviews. Thus the definition of each level was simplified and the number of levels was reduced to four. In addition to fluency, pronunciation, grammar, vocabulary, and comprehension are covered in Clark and Swinton's statements.

3) Harris proposes his set of definitions for measuring fluency. His other areas for evaluation include pronunciation, grammar, vocabulary, and comprehension.

Common among the three examples is the fact that temporal and sequential aspects of speech output are emphasized as central to the concept of fluency. Our definition of fluency is in accord with them. The above three examples are indicative of what aspects of speech output are important when native speakers evaluate foreign language learners' fluency.

However, there seem to be some problems in these techniques and definitions for evaluating fluency. First, the scoring tends to depend heavily on the rater's impression. This is due to practical requirements that the scoring procedure should be convenient and not time-consuming. Second, this dependence upon the rater's impressions necessitates that the definition of each level of fluency be succinct. However, the expressions used in the definitions are often ambiguous. What is meant by "uneven", "effortless", or "non-native" pauses? They may have some appeal to a native speaker's intuition, but if we hope to develop a valid measure of fluency which can be administered even by non-native speakers of English. we need to state the nature of fluency in more specific terms. If such an objective measure of fluency was developed, it would be of immense value to, particularly, non-native teachers or testers.

Thus it seems imperative to clarify what factors are crucially involved in the concept of fluency, especially when native speakers of English evaluate a learner's fluency. By identifying these factors, we can specify the nature of fluency and find a way to quantify the learner's fluency level. This goal is admittedly hard to achieve. A series of experiment were carried out as a small step toward this goal.

CHAPTER IV

EXPERIMENTS

So far we have defined the concept of fluency, considered various factors affecting it and surveyed current techniques of evaluating fluency. In this process the following two hypotheses were formulated:

> 1) the four factors (Speed of Response, Length and Amount of Hesitation Pauses, Number of Repetitions and False Starts, Logical Coherence and the use of Connectives) constitute a significant part of the overt evidence from which native speakers judge a foreign language learner's level of fluency,
>
> 2) fluency is characterized as integrative in nature if compared with other aspects of the speaking ability,

and one problem was identified:

> 3) impressionistic rating of fluency, which heavily depends upon native speakers' intuition, is quite prevalent in the evaluation of fluency.

As a small step toward the verification of the two hypotheses and the solution of the problem, each of them are further elaborated in the form of question and attempts will be made to answer these questions.

1. Questions asked

In connection with the hypothesis 1) above, the following question was asked:

1) To what degree do several factors (Words per Minute, Percentage of Hesitation Pauses in the total time of speaking, Mean Length of Hesitation Pauses and Number of Sentences per Minute) respectively influence native speakers' impressionistic evaluation of fluency?

All of them measure temporal or quantitative aspects. As for WPM, we would naturally expect this variable to correlate with native speakers' impressions of fluency. Up to a certain point beyond which communication becomes impaired, we could say that the faster one's speaking is, the more fluent he is ("fluent", in the narrow sense as defined earlier). It is also easy to imagine how the next variable, the percentage of the time spent in hesitation pauses in the total time of speaking, can be correlated with native speakers' impressions of fluency. The more one stumbles and hesitates, the less fluent he is. Mean Length of Hesitation Pauses represents a measurement of how long, on the average, the subject took in pausing during his speech. The last variable is closely related to the first variable, WPM, in that it measures the quantity of language units produced within a certain time.

Even though we defined fluency in a relatively narrow sense as a component in speaking skills mostly related to the quantitative aspects of those skills, it would , of course, be impossible to predict with precision native speakers' judgements of fluency solely from quantifiable measures such as the ones referred to in question 1) (Lehtonen 1978: 12).

Rather, our concern here is to verify the validity of the definition of fluency we offered above by choosing practically manageable measures of fluency and correlating them with native speakers' judgements of fluency. Further, we hope to determine which of the measures chosen in this study are reasonably accurate indices of fluency. This study makes no claim to be exhaustive or complete in highlighting these various factors operative in the concept of fluency.

Concerning the hypothesis 2), we will verify it by answering the following question:

> 2) Among the four aspects of speaking ability (Grammar, Vocabulary, Pronunciation, Fluency), which is the most integrative in nature?

In our consideration of the problem 3), we will focus on the hesitation pauses of non-native speakers. In an attempt to clarify what is meant by "non-native pauses", a term frequently used in evaluative definitions of fluency (e.g. Clark and Swinton 1979: 66), the following question needs to be asked:

> 3) If compared with native speakers' HP's, are there any differences in the duration and distribution of HP's in non-native speakers' English (i.e. the length and placement of HP's within a sentence)?

The underlying rationale here is that, if we can detect the differences between them, these aspects of HP's which are found to be different between native and non-native speakers would presumably characterize "non-native pauses", and

consequently be possible candidates for useful indices in fluency evaluation.

2. Method

Ten non-native speakers of English were asked to describe simply what happened in a comic strip (see Appendix I). The administration of the whole precedure was conducted by a female ESL teacher at the Intensive English Institute at the University of New Mexico. The comic strip was used, first to make sure that the content to be explained was immediately understandable to each subject, and second to standardize the content to be described so that the subjects would have to perform roughly the same sort of speaking task. The time of the subjects' description ranged from 36.4 seconds to 109.7 seconds. Each subject's description was recorded on a Sony reel-to-reel recorder Model TC106A, and used as the basic material in the following two experiments:

Experiment 1 (Native speakers' subjective evalution of non-native speakers' fluency and factors affecting it.)

Experiment 2 (Rate of speech and hesitation phenomena in non-native speakers' English as compared with those in native speakers' English.)

3. Subjects

Ten non-native speakers of English who were studying at the Intensive English Institute at the University of New Mexico served as subjects for this study. In order to secure optimum

variance in scores assigned to each subject, the subjects were so selected that their proficiency levels would range from the lowest to the highest represented at the Institute. Native language backgrounds included Spanish, Japanese, Persian, and Indonesian. Ages ranged from 18 to 31. There were 8 males and 2 females in the sample. Their length of stay in the United States ranged from 3 months to 2 years and their time in studying English from 3 months to 10 years.

4. Experiment 1

4.1

Purpose: The purpose of this experiment was to determine the degree of contribution of several factors, which can lend themselves to physical measurement, to the subjective evaluation of fluency by native speakers of English, consequently to answer questions 1) and 2).

4.2

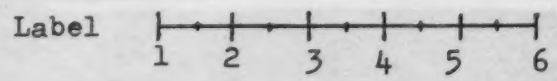
Subjective Evaluation by Native Speakers: The recordings of the ten subjects' spoken English were rated by 15 American native speakers of English. All of the raters were volunteers from courses in linguistics or the teaching of English as a second language at the University of New Mexico. 13 of them were graduate students and 2 undergraduates. Thus they could be said to be to some extent linguistically more

sophisticated than ordinary lay natives.

As this experiment utilizes a part of the data obtained in a previous study (Yorozuya and Oller 1980), the design of the experiment was not set up particularly for the present study. However, this proved to be an advantage rather than disadvantage for the present study, since thanks to the design it was possible to obtain additional interesting information as will be shown later.

The original recordings were copied five times onto a Sony CF370 cassette. After copying a whole series of the subjects' speech (No.1-10), the order of subjects was systematically rearranged to minimize recall of ratings assigned previously. In all, the raters listened to each subject's speech exactly 5 times. It took one hour and ten minutes to complete the whole listening and rating process. On each of four of the five listenings, ratings were assigned on only one of four scales (Grammar, Vocabulary, Pronunciation, Fluency). On one listening, ratings were assigned on all of the four scales simultaneously. In order to counterbalance a possible order effect, the raters were divided into five groups (A-E). Group A rated each subject's response on the first listening on all four scales, then on each of four successive listenings they rated the same ten responses (but in a different order each time) on Grammar, Vocabulary, Pronunciation, and Fluency, respectively. Group B rated all ten responses on the Grammar

scale first, then on all four scales, and so forth as shown on Table 1. Each scale thus appeared in each section (condition 1 = all scales rated at the same listening; condition 2 = each scale rated at a separate hearing). Each scale was rated exactly twice by each rater (read across the rows of Table I), but only once on any occation by any rater. The definition of scales including fluency was presupposed. This was because the primary concern in this experiment was to determine the degree of the contribution of several factors to the raters' overall impression of fluency, in an attempt to clarify the concept of fluency generally held by native speakers of English. The raters assigned ratings by putting a check on a 6 point scale with half-way points as shown below.



As a result, each subject's speech recieved 24 ratings on each listening, a total 120 ratings after five listenings. With reference to fluency, each subject's speech received three ratings on Condition 1 and three ratings on Condition 2 on each listening. After five listenings, the total number of ratings amounted to thirty, of which fifteen were rated on Condition 1 and fifteen on Condition 2.

TABLE 1

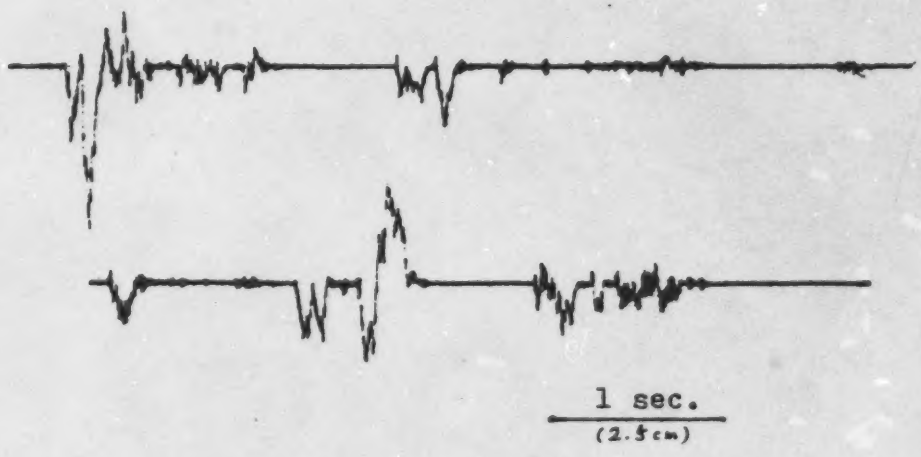
DESIGN MATRIX: ORDER OF SCALES RATED ON FIVE LISTENINGS BY FIVE GROUPS OF RATERS*

Group	Repeated listenings				
	1st	2nd	3rd	4th	5th
A (Raters 1-3)	All	Grammar	Vocabulary	Pronunciation	Fluency
B (Raters 4-6)	Fluency	All	Grammar	Vocabulary	Pronunciation
C (Raters 7-9)	Pronunciation	Fluency	All	Grammar	Vocabulary
D (Raters 10-12)	Vocabulary	Pronunciatio	Fluency	All	Grammar
E (Raters 13-15)	Grammar	Vocabulary	Pronunciation	Fluency	All

* On the first listening Group A rated each of the 10 subjects on all four scales, while Group B assigned ratings on all four scales on the second listenings and so forth.

4.3

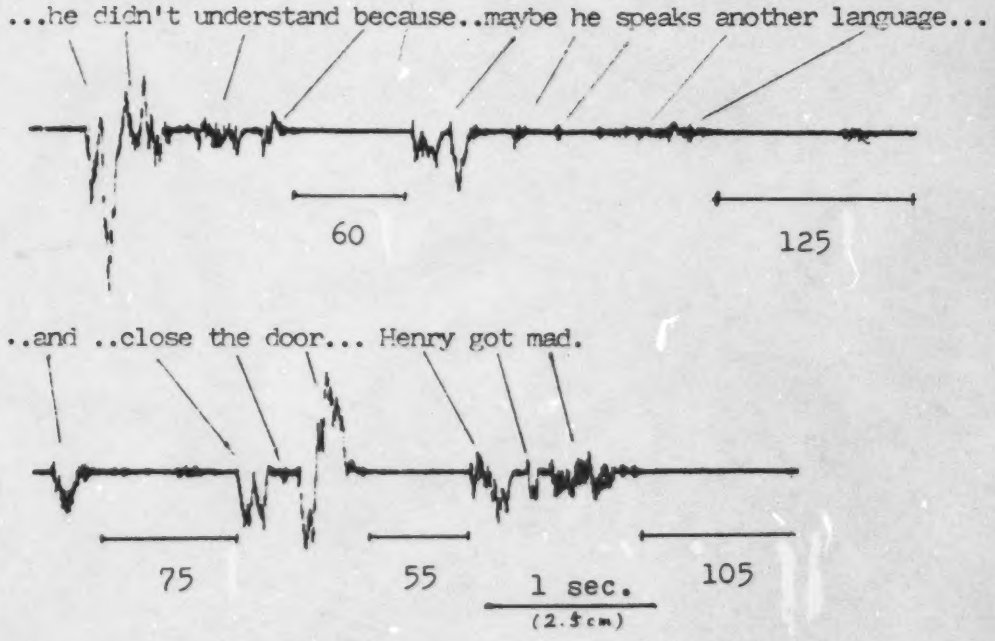
Transcription and Objective Measurement: The ten recordings of the non-native subjects' picture descriptions were fed into a pen-oscillograph in order to be transformed into a visual record. The use of a pen-oscillograph permitted a great increase in accuracy in measuring the temporal aspects of the subjects' speech, since measurement with a stopwatch, which tends to be unreliable and an extreamly tiresome process for the experimenter, was replaced by simple measurement with a ruler. A sample of the output from the equipment looks like this:



From this line, we can easily see where the subject was vocalizing and where he was not. One second elapsed time corresponds to 2.5 cm. The example above is 24 cm., which

means that it lasted for 9.6 seconds.

Next, while listening to the tape recording the speech was transcribed above the oscillograph line, taking care to match the waves with the corresponding words. Also, hesitation pauses longer than 0.5 seconds were located and identified by underlining and noting their duration in 1/100ths of 1 second. A specimen of the resulting text is shown below;



In all, the texts of the non-native subjects' speech totaled 15.71 metres in length, those of the native subjects 11.69 metres (used in the Experiment 2 below).

4.4

Basic Data: Before we consider the central findings of Experiment 1 to be shown in Table 6 and 7, it seems advisable to consider the basic data on which Table 6 and 7 are based.

Table 2 shows the ten non-native speakers' total time of description, number of words used in the description, Words per Minute (WPM). Total description Time indicates the time spent in describing the series of pictures, i.e. from the first word to the last word used in the description. The shortest total time was 36.4 seconds (Subject 2), the longest 109.7 seconds (Subject 6), and the mean was 62.84 seconds. The Number of Words was counted excluding repetitions of the same phrase or word, false starts, and unintelligible words (e.g. "... and and his friend ... well Henry ask for something." Words with x below were not counted). The lowest number was 43 words (Subject 8), the highest 114 words (Subject 5), and the mean was 78.4 words. WPM was computed by dividing the Number of Words by the Total Description Time and multiplying the results by 60. The lowest rate was 49 words per minute, the highest 99.7, and the mean 74.9.

Just in passing, we could compare the above results with the native speakers' performance in Experiment 2 described in the next section, in order to see how close (or far) the ten non-native speakers in the present study are to the native speakers' norm.

TABLE 2

TOTAL DESCRIPTION TIME, NUMBER OF WORDS, AND WORDS PER MINUTE OF THE SUBJECTS' RESPONSES

(NON-NATIVE SPEAKERS)

Subject	Total Description Time*	Number of Words	Words per Minute (WPM)
1	46.6	72	92.7
2	36.4	59	97.3
3	65.6	109	99.7
4	40.7	63	92.9
5	90.0	114	76.0
6	109.7	87	47.6
7	69.0	57	49.6
8	40.8	43	63.2
9	78.0	102	78.5
10	51.6	78	90.7
Total	628.4	784	
Mean	62.84	78.4	74.9

* The figures are in seconds.

TABLE 3

TOTAL DESCRIPTION TIME, NUMBER OF WORDS, AND WORDS PER MINUTE OF THE SUBJECTS' RESPONSES

(NATIVE SPEAKERS)

Subject	Total Description Time *	Number of Words	Words per Minute (WPM)
1	48.0	106	132.5
2	30.8	93	181.2
3	120.8	283	140.6
4	25.8	81	188.4
5	48.6	102	125.9
6	38.1	108	170.1
7	25.9	50	115.8
8	34.4	87	151.7
9	74.3	169	136.5
10	21.2	73	206.6
Total	467.9	1152	
Mean	46.79	115.2	147.7

*These figures are in seconds.

Table 3 shows the ten native speakers' performance on the three dimensions. The mean of the Total Description Time was 46.79 seconds, which means that the native speakers, on the average, did not require so much time as the non-natives, though there is substantial overlap between the two groups. The mean of the Number of Words of the native speakers was 115.2, which is larger than the non-natives' mean, 78.4. Most remarkable is the native speakers' WPM mean of 147.7, nearly double that of the non-natives' mean, 74.9. If we assume that each word conveys about the same amount of information, the natives were able to convey twice as much information as the non-natives were able to, in describing the comic strip in the present study. Thus, as far as this study is concerned, we could say the native speakers' efficiency in handling the oral production of English was about twice as high as that of the non-native speakers.

Table 4 shows various measurements concerning the sentences used by the non-native subjects. The average number of sentences used was 10.4. The number of words in a sentence, that is the average length of sentences was 7.5 words. And the average number of sentences per minute was 9.9. Though, in fact, there were several cases where it was doubtful whether the utterance should be counted as a sentence, the following rationale was observed;

TABLE 4
NUMBER OF SENTENCES, NUMBER OF WORDS PER SENTENCES, AND NUMBER OF SENTENCES PER MINUTE, OF THE NON-NATIVE SPEAKERS' SPEECH

Subject	Number of Sentences	Number of Words per Sentence	Number of Sentences per Minute
1	10	7.2	12.9
2	10	5.9	16.5
3	11	9.9	10.1
4	9	7.0	13.3
5	15	7.6	10.0
6	17	5.1	9.3
7	6	9.5	5.2
8	6	7.1	8.8
9	10	10.2	7.7
10	10	7.8	11.6
Mean	10.4	7.5	9.9

"Shall I do your work?" → one sentence.

"Mr. Smith is angry and close the door..." → two sentences.

"Henry came and see him." → one sentence.

"He didn't understand, because maybe he spoke another language." → two sentences.

"... and he did ... his front door's messy." → one sentence (a false start).

"He said to Mr. Smith that the yard is clean." → two sentences.

"Henry angry ... was angry." → one sentence (self-correction.)

Hesitation pauses were counted and measured as described in 4.3.

Concerning the hesitation pauses dealt with in the present study, two points should be mentioned. First, the distinction between juncture and hesitation pauses has been disregarded. Some researchers point out it is possible to clearly distinguish the two types of pauses. For instance, Maclay and Osgood (1959) state as follows:

> there is one basic difference between pause and hesitation: hesitations, as we define them, refer to events that are relatively gross and easily observable, while the juncture pauses of linguistic theory are quite short in duration and much harder to observe and record. In addition, hesitations often interupt the flow of speech while juncture pauses do not.

Through analysing the texts in this study, however, it became immediately apparent that this distinction is quite fuzzy. In other words, it was difficult to distinguish between juncture pauses and hesitation pauses. There were some cases of pauses which sounded like hesitation pauses (e.g. a long silence with [ə:] or [ər] sound) even though, judging from their location in the sentence, they could be analyzed as juncture pauses. As a result of this conceptual imprecision, there was considerable overlap between the two types of pauses. Thus, the distinction was disregarded, and pauses dealt with

in this study therefore include both types. The term "Hesitation Pauses" used in this study thus includes both juncture and hesitation pauses as defined by Maclay and Osgood (1959). Second, the threshold level beyond which all silent periods in speech were regarded as hesitation pauses was set at 0.5 seconds. This was decided somewhat arbitrarily by intuition through a number of actual listenings to the recorded speech of the subjects. It was felt that pauses longer than 0.5 seconds sounded like hesitation. To corroborate our claim that this does not deviate greatly from the native speakers' intuitions, we can adduce support from Boomer and Dittmann (1962)'s finding that, in cases of hesitation pause, any pause exceeding 0.5 second is almost sure to be perceived as a hesitation; in cases of syntactically determinable pause i.e. juncture pause, they are perceived as pauses if they exceed 0.5 - 1 second.

Thus, it can be affirmed that a threshold level of 0.5 seconds is quite suitable for hesitation pauses, but less so for juncture pauses. We could have set the level at one second or more above which probably most of the pauses at any position would be considered as grave or halting. However, if we did so, we would ignore many hesitation pauses at syntactically unpredictable points which are shorter than the level. Thus, no matter which level is chosen, we have to allow for some

compromise, as we have determined to treat juncture and hesitation pause indiscriminately. Consquently, with the likely possibility in mind that we might, by setting the threshold level at 0.5 seconds indiscriminately for both types of pauses, inadvertently include in the statistics some juncture pauses not usually judged to be halting, we will apply this level which may be presumed a reasonable approximation.

In short, we consider any pauses exceeding 0.5 seconds as hesitation pauses regardless of when and where they occur in the subjects' speech.

Some basic data concerning the hesitation pauses observed in the non-native subjects' speech are shown in Table 5. Total Description Time (in the first column) is identical with the second column of Table 2 and is reproduced for convenience. In the second column, we can see how long the subjects remained silent in their speech (including er's and erm's). The range is 14.85 - 63.8 seconds, and the mean is 31.515 seconds. It should be remembered that all the silent periods in speech were not included in the computation of Total Hesitation Time but only those beyond 0.5 seconds. Number of Hesitations in the third column is the number of units of silence during which there was no intervention of a meaningful word except for er's and erm's. The range is 17 - 43, and the mean is 25.6. This variable alone does not indicate accurately the gravity of

TABLE 5

HESITATION PAUSES OF THE NON-NATIVE SPEAKERS

Subjects	A. Total Description Time (seconds)	B. Total Hesitation Time (seconds)	Number of Hesitations	Mean Hesitation Length (seconds)	B / A
1	46.6	17.65	19	0.93	37.9%
2	36.4	14.85	20	0.74	40.8%
3	65.6	28.80	22	1.31	43.9%
4	40.7	16.35	17	0.96	40.2%
5	90.0	45.75	43	1.06	50.8%
6	109.7	63.80	41	1.56	58.2%
7	69.0	40.35	26	1.55	58.5%
8	40.8	26.85	17	1.58	65.8%
9	78.0	39.70	34	1.17	50.9%
10	51.6	21.05	17	1.24	40.8%
Total	628.4	315.15	256		
Mean	62.84	31.515	25.6	1.23	50.2%

* All the hesitation pauses measured in the present study are over 0.5 second.

hesitations in the subjects' speech. Mean Hesitation Length shows how long each subject remained silent per unit of hesitation pause. This mean was thus calculated by dividing Total Hestitation Time by Number of Hesiations. The range is 0.74 - 1.58 second, the mean 1.23 second. The last column shows the percentage of Total Hesitation Time in Total Description Time. The range is 37.9% - 65.8%, the mean 50.2%. This means about 50% of the subjects' speech was occupied by hesitation pauses longer than 0.5 seconds. If all the pauses shorter than 0.5 seconds were included, the percentage would be much higher.

4.5

Central findings: Having examined the basic data, we can now proceed to consider possible answers to Questions 1) and 2), the central concern in Experiment 1.

Concerning the first question about possible factors contributing to native speakers' impression of fluency, we have chosen four factors: WPM, Percentage of Hesitation Pauses, Mean Hesitation Length, Number of Sentences per Minute (they can be found in Table 2, 5, 4, 5, respectively). These four variables were correlated with the fifteen native speakers' subjective evaluations of fluency. As was explained in 4.1, there were two conditions under which the native speakers assigned ratings on fluency: Condition 1 and Condition 2.

Under Condition 1,fluency was rated at one listening together with the remaining three scales: Grammar, Vocabulary and Pronunciation. Under Condition 2, fluency only was rated at one listening. So it is reasonable to assume that the native speakers concentrated on fluency more under Condition 2 than 1.

The reliability of the fifteen native speakers could be estimated from the correlation between means of their ratings assigned under Condition 1 and those under Condition 2. We could take this correlation as a kind of test-retest reliability coefficient. The result was .95 which implies the fifteen native speakers as a whole can assign quite consistent ratings on different occasions.

Referring to Table 6, we can discover which of the factors played an important role when native speakers of English formulated their impressions of fluency while listening to the ten subjects' speech.

First, let us consider which of the two dependent variables ratings under Condition 1 or under 2 is a more valid measure of fluency. All of the four variables, correlated better with the ratings under Condition 2 than with the same ratings under Condition 1. As was mentioned earlier, the four variables were chosen because they were considered to be likely factors in objective measurements of fluency. Thus, it would be fair to say that a variable which correlates better with them would

TABLE 6

CORRELATIONS OF NATIVE SPEAKERS' SUBJECTIVE EVALUATION OF FLUENCY (CONDITION 1, CONDITION 2) WITH OBJECTIVE MEASURES OF FLUENCY

	Variables	Native speakers' subjective evaluation of fluency	
		Condition 1	Condition 2
1	Words per Minute	.79**	.85**
2	Percentage of Hesitation Pauses	-.65*	-.78**
3	Mean Hesitation Length	-.36	-.57
4	Number of Sentences per Minute	.36	.53

* $P < .05$, ** $P < .01$.

be a more valid measure. For this reason, we focus on the ratings under Condition 2. This is primarily due to the fact that under Condition 2 the native speakers concentrated solely on fluency while their attention to fluency was, so to speak, distracted under Condition 1 where they were not able to distinguish it from the other three scales: Grammar, Vocabulary and Pronunciation. This kind of judgemental bias is called a "halo effect" which refers to raters' tendency

to assign an inadequately high or low rating to one aspect which has been affected by their impression of the other aspects.

Among the four variables, Words per Minute is apparently the best predictor of native speakers' evaluation of fluency, the correlation being .85 which is significant at .01 level. This positive correlation means that those subjects whose speed of utterance was great tended to the rated highly on the fluency scale. The second best predictor was the second variable, Percentage of Hesitation Pauses, the correlation coefficient being -.78 which is significant at .01 level. This means that the longer the subjects remained silent, i.e. hesitated during their speech, the lower ratings they tended to be assigned. The other variables did not reach .05 significance level. The variables 3 and 4 have about the same relatively low levels of correlation.

According to Connonlly and Sluckin (1971: 154) [4], WPM and Percentage of Hesitation Pauses have a "high" correlation or "marked relationship" with the native speakers' fluency rating under Condition 2, and Mean Hesitation Length and Number of Sentences per Minute have a "moderate" correlation or "substantial relationship" with that fluency rating.

If we interpret the correlation coefficients in terms of percentage, in the case of WPM, we can say that 72% ($=.85^2$) of the variance of the ratings assigned could be predicted by scores on WPM. Accordingly, in the case of Percentage of

Hesitation Pauses, 61% ($-.78^2$) of the variance of fluency can be predicted.

Question 2) is concerned with which of the four aspects of speaking ability, i.e. Grammar, Vocabulary, Pronunciation, Fluency, is the most integrative, when these are evaluated impressionistically by native speakers. From the limited data available to us we might attempt to answer this question by seeing how much each of the four scales correlates with the remaining three. Table 7 shows the intercorrelations among the four scales under Conditions 1 and 2. Referring to the first column, for instance, the correlation between Grammar scale under Condition 1 and Vocabulary scale under Condition 1 is .89. The correlation between Grammar scale under Condition 1 and Pronunciation scale under Condition 1 is .85. Then, going to the lower column, the correlation between Grammar scale under Condition 1 and Grammar scale under Condition 2 is .87, and so on. Mean correlation indicates how much on the average each scale correlated with the other remaining scales. For example, in the last column, we can see that Fluency scale under Condition 2 had an average correlation of .89 with Grammar, Vocabulary, Pronunciation, Fluency scales under Condition 1, and an average correlation of .79 with Grammar, Vocabulary, Pronunciation scales under Condition 2. Arranging the correlations into four groups (.87, .89, .84, .89 / .80, .76, .84, .89 / .79, .87, .80, .84 / .76, .74, .72, .79), we

TABLE 7

CORRELATIONS ACROSS THE FOUR SCALES UNDER THE TWO CONDITIONS

Scales	Grammar 1	Ponunciation 1		Grammar 2	Pronunciation 2		
		Vocabulary 1	Fluency 1		Vocabulary 2		Fluency 2
Grammar 1	1.00	.89	.85	.88	.87	.61	.85	.82
Vocabulary 1	.89	1.00	.82	.96	.82	.85	.79	.99
Pronunciation 1	.85	.82	1.00	.84	.74	.70	.95	.78
Fluency 1	.88	.96	.84	1.00	.77	.86	.76	.95
Mean Correlation *	.87	.89	.84	.89	.80	.76	.84	.89

Scales	Grammar 1	Pronunciation 1		Grammar 2	Pronunciation 2		
		Vocabulary 1	Fluency 1		Vocabulary 2		Fluency 2
Grammar 2	.87	.82	.74	.77	1.00	.70	.80	.76
Vocabulary 2	.61	.85	.70	.86	.70	1.00	.61	.87
Pronunciation 2	.85	.79	.95	.76	.80	.61	1.00	.73
Fluency 2	.82	.99	.78	.95	.76	.87	.73	1.00
Mean Correlation	.79	.87	.80	.84	.76	.74	.72	.79

*Mean Correlations were computed by adding the three or four correlations in a row. For instance, the first Mean Correlation (.87) is the mean of .89, .85, .88. Since simple addition of correlations would bias the result, the following formula was used:

$$\text{Mean Correlation} = \sqrt{\frac{(r_a)^2+(r_b)^2+(r_c)^2\left(+(r_d)^2\right)}{3(\text{or } 4)}}$$

(Welkowitz, et. al.1976)

can see that in three cases out of the four, Fluency scale has the highest correlations with the other scales. As all the four scales showed a fairly high correlaion with each other, we could say that all the four aspects are more or less integrative. But still it is noteworthy that the Fluency scale is the most integrative of the four, at least within the native speakers' conceptual framework. This finding is in accord with our supposition voiced in Chapter II, section 3 that the concept of fluency is closely related to other aspects of non-native speakers' ability to speak, though it is emphasized that the concept of fluency is primarily concerned with the temporal and sequential aspects of speech output.

5. Experiment 2

5.1

Purpose: In order to determine what aspects of hesitation pauses characterize "non-native pauses" and thus to answer question 3), another experiment utilizing the data from Experiment 1 was conducted. It is hoped that if we can clarify what is meant by "non-native pauses", a term which appears often in definitions of fluency, the notion of fluency itself would also become more explicit.

5.2

Method: The method adopted here is comparative. By comparing non-native speakers' hesitation pauses with those of

native speakers, we will be able to clarify what characterizes "non-native pauses". Additionally, ten native speakers of English (Americans) currently living in Hiroshima were asked to describe the same comic strip. Their age ranged from 20 to 51 and the time spent in Japan 3 months to 3.5 years. Their tape recordings were transcribed and hesitation pauses were located and measured.

5.3

Findings: First let us look at Table 8, which mainly shows the differences in temporal aspects of hesitation pauses between the two groups. Total Description Time is the average time spent by each subject in explaining the comic strip. Total Hesitation Time is the average of the total time spent by the ten subjects in each group.

First let us compare Total Hesitation Time. The figures 15.54 and 31.515 should not be compared directly since they do not have the same basis. We have to convert these into comparable figures by calculating an "estimated time of

TABLE 8
A COMPARISON OF HESITATION PAUSES BETWEEN NATIVE AND NON-NATIVE SPEAKERS

Group	A. Total Description Time	B. Total Hesitation Time	C. Number of Hesitation	Mean* Hesitation Length	B/A
Native	46.79	15.54	16.3	0.95	33.2%
Non-Native	62.84	31.515	25.6	1.23	50.2%

All the figures in the first three columns are in seconds.
*Computed by dividing B by C.

hesitation pauses per minute". Thus 15.54 seconds becomes 19.93 and 31.514 seconds becomes 30.09. The two converted figures differ by 10.16 seconds. This indicates that each of the non-native subjects paused on the average 10.06 seconds per minute longer than the native subjects did. Number of Hesitations should also be converted to allow for a direct comparison. The resulting figures are 20.9 for the native subjects and 24.44 for the non-native subjects. It follows that the non-native subjects paused on the average almost four more times per minute than the native subjects. Figures in the next column, Mean Hesitation Length allows a direct comparison. The figures indicate that, on the average, the native subjects paused 0.95 seconds for one hesitation pause, while the non-native subjects hesitated for 1.23 seconds(significant at .001 level). The percentage of hesitation pauses in the total time spent in description also differs greatly. For the native subjects, about one third of the total time is occupied by hesitation pauses longer than 0.5 seconds, for the non-native subjects about a half. On the whole, the non-native subjects could be said to have used significantly more hesitation pauses.

To summarize the above findings, the non-native subjects paused more frequently and longer than the native subjects. Thus the implication we can extract from this finding for the problem of what characterizes "non-native pauses" is that non-native pauses are probably much longer in duration and

more frequent than those of native speakers. As this is a very broad statement, the following analyses attempt to investigate the finer differences between these two kinds of hesitation pauses.

Our next analysis is concerned with the frequency of hesitation pauses of various length. Figure 5 shows the result. The broken line represents the number of the native subjects' hesitation pauses observed in one minute and the solid line represents the non-native subjects' pauses. The interval in the regular frequency polygon was set at 0.27 seconds, which was determined to allow for the best pictorial representation of the general tendencies of the two groups. The means of the two groups are indicated by △ (Native subjects = 0.95 seconds) and ○ (Non-native subjects = 1.23 seconds). It should be noted that the initial drops in the two lines at 0.5 seconds are due to the fact that, in the first interval (0.365 - 0.635 seconds), only hesitation pauses with the length of 0.5 - 0.635 seconds were counted. Thus we counted only about half of the hesitation pauses in this interval. If we doubled the hight of the initial points, we would see the lines begin from much higher points.

First, concerning the distributional patterns, the experimenter's initial expectation was that the two lines would show quite different shapes and the peaks would be located at different places. However, the result turned out to be an unexpected one. An interesting fact we can observe in the

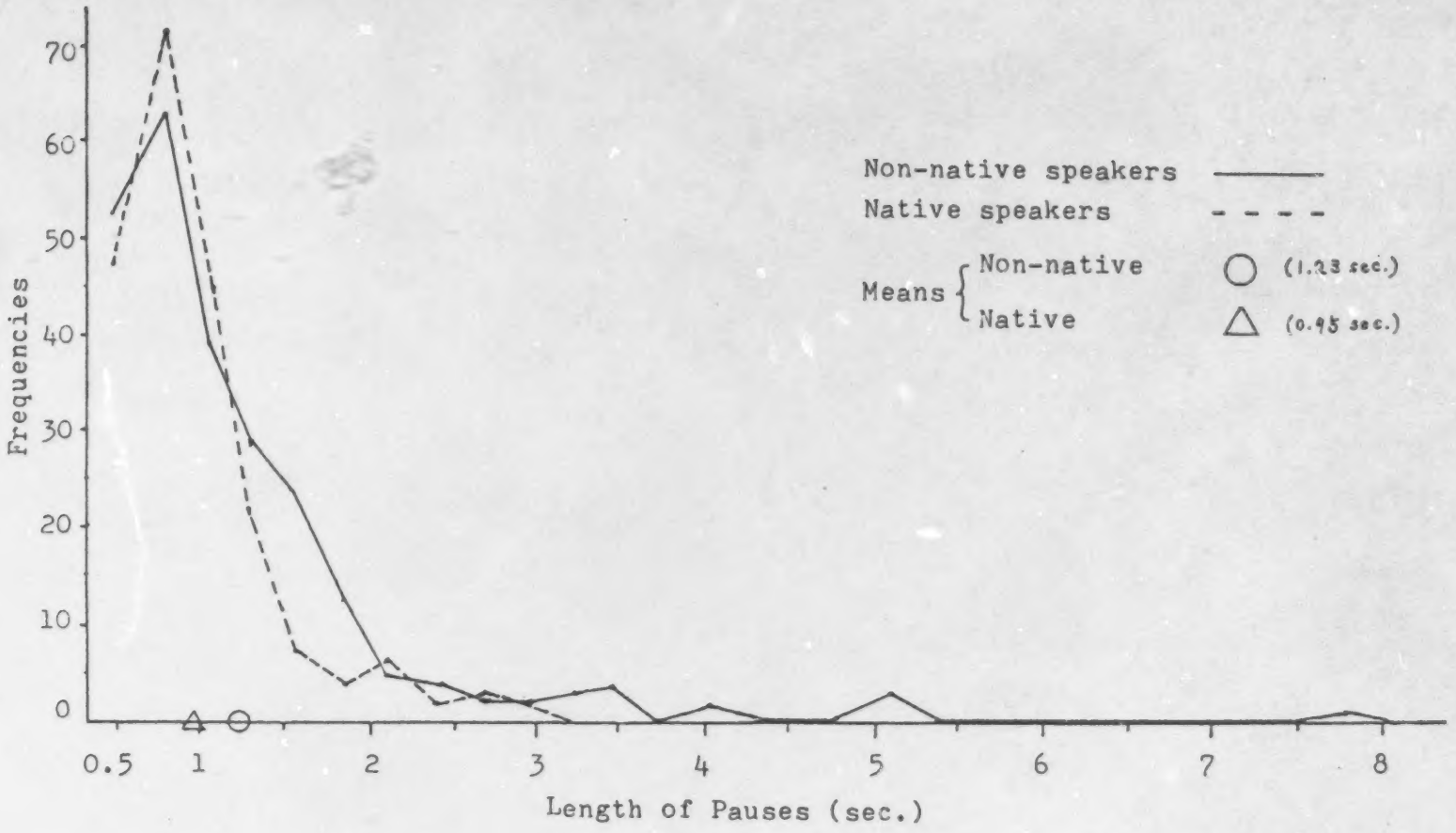


FIGURE 5
FREQUENCY DISTRIBUTION OF HESITATION PAUSES IN NON-NATIVE AND NATIVE SPEAKERS' SPOKEN ENGLISH

figure is that the two lines show, though not identical, quite similar shapes. This means that in general non-native speakers do not differ from native speakers in the distribution of frequencies of hesitation pauses with various lengths.

However, we can notice some differences between the two groups, which might offer us hints for the question "What is meant by 'non-native pauses'?". The greatest difference between the two lines is in the interval from 1.445 - 1.715 seconds. There were twenty-four hesitation pauses in one minute for the non-native subjects. The difference is sixteen. Another noticeable difference is that there are no hesitation pauses of the native speakers longer than 3 seconds, whereas there are thirteen such hesitation pauses in the non-native speakers' data. Though nothing can be inferred about hesitation pauses shorter than 0.5 seconds, it would seem that the main differences between the two lines are found around 1.5 seconds and above 3 seconds.

An interesting fact is that the native subjects exceeded the non-native subjects at four points of the lines: at the intervals 0.635 - 0.905 sec., 0.905 - 1.175 sec., 1.985 - 2.255 sec. and 2.525 - 2.795 sec. the latter two intervals, the difference may be due to a sampling error, i.e. the fact that the sample was too small. In the first interval where the peaks of the lines fall, the difference is nine (Natives = 72, Non-natives = 63). In the second interval, it is six (Natives = 45, Non-natives =39). These differences could also

be due to a sampling error. But it is more likely that native speakers as a whole tended to have more hesitation pauses whose lengths fall between 0.635 - 1.175 seconds.

Admitting that our data is quite limited in size, a general finding could be that, the native subjects tended to have slightly more hesitation pauses with the lengths of 0.635 - 1.175 seconds while the non-native speakers tended to have more and longer hesitation pauses, i.e. those exceeding 1.5 seconds. This in turn suggests that a non-native speaker of English sounds "non-native" because he or she has many hesitation pauses longer than about 1.5 seconds, and in addition that he or she can have quite a lot of hesitation pauses whithout sounding "non-native" if they are 0.635-1.175 seconds in length. These generalizations are necessarily tentative and speculative. The latter especially is quite uncertain, while the former seems to appeal to common sense.

The following limitation of the present analysis should be recognized: we have neglected the study of the locations of the hesitation pauses, that is precisely where the hesitation pauses occurred within an utterance. As was noted earlier it has been shown by Boomer and Dittmann(1962) that the threshold levels for juncture pauses and hesitation pauses are different. Thus, it is quite conceivable that a non-native speaker who paused longer than 1.5 seconds would still sound "native-like", if the pause occurred between sen-

tences. If a pause of the same length occurred between, say, an adjective and a noun (e.g. much...snow), it would presumably sound "non-native". Therefore, our speculation above that a non-native speaker sounds "non-native" because he or she has many hesitation pauses longer than about 1.5 seconds is valid only in broader terms and needs elaboration, with which we will next be concerned.

So far we have been concerned with temporal aspects of hesitation pauses without regard to their locations in an utterance. This is obviously insufficient. We will next consider the lengths and frequencies of hesitation pauses in relation to their locations. Then, we will compare the two groups in this regard in order to investigate further into the problem of what characterizes "non-native pauses".

First, we need to set up a descriptive framework for revealing the difference between the two groups in distribution of hesitation pauses within a sentence. There are a few previous research studies which can offer such a framework. Boomer and Dittmann(1962) used a "phonemic clause" (a term which is usually referred to as a tone group or a breath group) and analysed hesitation pauses in terms of their positions in a phonemic clause. This framework was not employed because of a weakness, namely, it was difficult to determine where a phonemic clause began and ended. Consequently, the framework by Goldman Eisler (1968) was adopted on the ground that it

uses the traditional notion of sentence. Her categories for describing hesitation pauses are as follows:

Grammatical Junctures

(1) "Natural" punctuation points, e.g. the end of a sentence.

(2) Immediately preceding a conjunction whether (i) coordinaiong, e.g. and, but, neither, therefore, or (ii) subordinating, e.g. if, when, while, as, because.

(3) Before relative and interrogative pronouns, e.g. who, which, what, why, whose.

(4) When a question is indirect or implied, e.g. "I don't know whether I will".

(5) Before all adverbial clauses of time (when), manner (how), and place (where).

(6) When complete parenthetical references are made, e.g. "You can tell that the words-this is the phonetician speaking-the words are not sincere".

Non-Grammatical Breaks

(7) Where a gap occurs in the middle or at the end of a phrase, e.g. "In each of//the cells of the body//..."

(8) Where a gap occurs between words and phrases repeated, e.g. (i) "The question of the//of the economy". (ii) This attitude is narrower than that//than that of many South Africans".

(9) Where a gap occurs in the middle of a verbal compound, e.g. "We have//taken issue with them and they are// resolved to oppose us".

(10) Where the structure of a sentence was disrupted by a reconsideration or a false start, e.g. "I think the problem of de Gaulle is the//what we have to remember about France is..."

Using these ten categories, the observed hesitation pauses were classified. However, as is always the case in utilizing someone else's descriptive framework for one's own study, the categories did not fit the present data. There were remarkably many examples which were not readily classified according to Goldman Eisler's categories. Consequently, an extensive revision was necessary and the number of categories was increased from ten to twenty four. Thus, the resultant categories together with Goldman Eisler (1968)'s ten categories are as follows:

1-6 = Hesitation pauses at Grammatical Junctures in Goldman-Eisler (1968)

7-10= Hesitation pauses at Non-Grammatical Breaks in Goldman-Eisler (1968)

11, Before an adverb or an adverbial phrase. e.g. "There's a lot of snow//in front of the door."

12, Before an infinitive "to". e.g. "... he goes to his neighbor's house//to see if... ."

13, After a subordinating or co-ordinating conjunction. e.g. "And//the neighbor agrees." "However//the owner of the next neighborhood didn't mean that."

14, After the verb "say" or "tell + pronoun" and before a reported speech. e.g. "Mr. Smith says//"Yes, you could do it."

15, Before a participle construction. e.g. "He stands at one side//sweat rolling down..."

16, After the subject and before a pronoun referring to it. e.g. "This guy//he rings the doorbell."

17, After the conjunction "that". e.g. "...what is happenning is that//he goes to a neighbor's house..."

18, After an adverbial phrase. e.g. "After that//he rest."

19, Between the subject and a verb. e.g. "Henry//is digging snow..."

20, After an article or an adjective and before a noun. e.g. "...obviously showing him the//job to be done..."

21, After a verb and before its object. e.g. "Henry shovels//the snow."

22, After a noun and before "of". e.g. "...the entrance //of the house..."

23, After a verb and before a complement. e.g. "Henry looks//disappointed."

24, After an infinitive "to" and before a verb. e.g. "He wants to//work."

The results of the analysis are shown in Table 9. For each category, mean length ($\bar{x}$) in seconds and the total number (n) of hesitation pauses observed in one minute are shown. We can see from the table how many hesitation pauses of the non-native subjects and the native subjects occurred at a certain point and how long they were on the average.

Since in many categories there were not sufficient examples in the present sample, categories in which either non-native or native subjects' hesitation pauses did not reach four in number were left out of consideration. It is dangerous to infer anything from a comparison of isolated examples. In addition, if more than seven either non-native or native subjects did not show hesitation pauses in a category, this category was left out of consideration in order to avoid a strong bias in favor of the idiosyncracy of a few individual subjects. An extreme case could occur where only one subject of the non-native group had, say, forty

TABLE 9

LENGTH AND NUMBER OF HESITATION PAUSES IN RELATION TO THEIR LOCATION

Location	1		2		3		4		5		6		7	
	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$	n
Non-natives	0.99	42	1.00	13	0	0	0	0	0	0	0	0	0.78	9
Natives	0.95	63	0.77	22	0.60	3	0	0	0	0	0	0	1.10	8

Location	8		9		10		11		12		13		14	
	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$	n
Non-natives	1.24	38	0.77	3	2.10	16	0.74	8	0.65	1	1.25	43	1.65	1
Natives	0.80	9	0.60	3	1.00	6	0.85	6	1.10	3	0.83	15	1.40	5

Location	15		16		17		18		19		20		21	
	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$	n
Non-natives	0	0	1.15	1	1.63	2	1.14	4	1.29	25	0.99	14	1.26	10
Natives	1.07	4	1.65	3	1.70	1	0.93	4	0.97	17	1.10	22	0.79	9

Location	22		23		24	
	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$	n
Non-native	0.58	2	1.06	5	2.60	3
Natives	0.65	1	1.23	5	1.00	1

hesitation pauses in Category 3, while each of the ten native subjects had three hesitation pauses in the same category. In such a case, the result turns out to be that on the average the non-native group had more hesitation pauses in that category, but this inference is greatly biased by one subjects' idiosyncracy.

Among Categories 1 - 6, which show the average lengths and number of hesitation pauses per minute observed at "Grammatical Junctures" (Goldman Eisler 1968: 13), only Categories 1 and 2 satisfied the two conditions mentioned above. Hesitation pauses in Category 1 occurred at "Natural" punctuation points, mostly, the end of a sentence, and those in Category 2 before co-ordinating or subordinating conjunctions. The result shows in those categories, the native subjects tended to have more and shorter hesitation pauses. This means the native subjects paused more frequently but for a slightly shorter time than the non-native subjects at the end of a sentence or before a conjunction. This should not be taken to mean that the native subjects were less fluent than the non-native subjects. The greater frequencies of the native subjects' hesitation pauses are due to the fact that they produced more sentences within a minute than the non-native subjects did. The next four categories 3 - 6 did not have great frequencies. This is probably due to differences between our study and Goldman Eisler's in the

kind of speech data to which these categories were applied. Our data consists of the recorded speech used in picture description whereas Goldman Eisler's consists of interviews.

Categories 7 - 10 show the data of hesitation pauses at "Non-Grammatical Breaks". A remarkable difference between the two groups is found in Categories 8 and 10. Hesitation pauses in Category 8 occurred between repeated words or phrases. The non-native subjects paused between repeated words or phrases about four times more frequently as well as for a longer time than the native subjects. Of the ten non-native subjects nine showed this kind of hesitation pause. It was found that in most of the cases, the non-native subjects repeated the first word of a determiner - noun, a subject - verb, or a verb - object sequence, while they tried to remember the second word of these sequences. This phenomenon has been called "prospective repeat" by Hieke (1981: 152). An example of this type of hesitation pause between repeated words is as follows;

> . . . Henry is cleaning the /2.2/ the snow in the /0.7/ in the walkway . . .
>
> (// = a hesitation pause with its duration in seconds)

This shows the non-native subjects had difficulty in remembering the word "snow" and "walkway". These repetitions with an intervening hesitation pause seem to function as a convenient device for non-native speakers of English to maintain

the continuity of speech when they have difficulty with vocabulary selection.

Hesitation pauses in Category 10 occurred when the subjects committed a false start or changed the sentence structure in the middle. The non-native subjects had about three times as many hesitation pauses of this type with average durations twice as long as those of the native subjects. This suggests that the non-native subjects practiced much more repairing of structures and individual words.

Categories 11 - 24 have been added to Goldman Eisler's ten categories. Categories 13, 19, 20, 21 satisfied the two conditions mentioned above.

Hesitation pauses in Category 13 showed statistically the greatest discrepancy between the two groups in mean length ($\bar{x}$) and this was the only category where the two means were different at a .05 significance level. This type of hesitation pause occurred immediately after conjunctions such as "and", "so", "but", and so on, or an adverb "then". These connecting words have been shown to be one of the most common features of native speakers' spoken English (Crystal and Davy 1975: 88). However, it seems reasonable to conclude that non-native speakers of English tend to pause much more frequntly and for a longer time immediately after those connecting words than native speakers. This seems primarily due to inefficiency in encoding.

The next great difference in hesitation pauses between the two groupscan be found in Category 19. The result shows that the non-native subjects' hesitation pauses were greater in length and frequency than the native subjects'. This indicates that the non-native subjects hesitated more after they uttered the subject of a sentence. In most cases they seemed to be struggling to find the proper verb with which to continue the sentence.

Category 20 reveals an interesting fact that the native subjects, contrary to the results in the other categories excepting 1, evidenced a greater amount of hesitation pauses between a determiner and a noun. Typical examples from the native subjects are as follows:

> . . . He's moving all the snow back onto the /2.1/ front walk.
>
> . . . and then he asks for his /1.0/ renumeration . . .

Most of the determiners were "he" or "his". Hesitation pauses after these words seem to be to gain time while searching for an appropriate noun to follow. It could be said that this type of hesitation pause is quite common in the native speaker's speech and is probably not a useful index of "non-native pauses".

The last category, 21, shows a greater discrepancy in the factor of length than in frequency. This means that the non-native subjects paused longer after a verb and

and before its object. Again the difference in richness of vocabulary seems to be the key differentiating factor.

To summarize the above findings, the length and the frequency of the hesitation pauses in Category 8 (between repetitions of words or phrases), Category 10 (when a false start is made), Category 13 (after connecting words), Category 19 (between the subject and a verb), Category 21 (between a verb and its object) were shown to be possible candidates for the characterization "non-native pauses".

6. Summary of the Findings

Experiments 1 and 2 were conducted in answer to Question 1) - 3). The results are summarized below.

The first question was answered in Experiment 1. The finding was that the native speakers' impressions of fluency correlated highly with WPM and Percentage of Hesitation Pauses, and that there was a moderate correlation with Mean Hesitation Length and Number of Sentences per Minute.

In answer to Question 2) , fluency was found to be the most integrative of the four aspects of speaking ability, having a high correlation with Grammar, Vocabulary and Pronunciation.

We attempted to answer Question 3) in Experiment 2. The findings were, first, that the non-native subjects

hesitation pauses were, as a whole, much longer in duration and more frequent than those of the native subjects. Secondly, there was not a substantial gap between the patterns of distribution of the hesitation pauses for the two groups. However, the non-native subjects tended to have more hesitation pauses longer than 1.5 seconds. Thirdly, greater differences were found between the native and the non-native subjects in the length and the frequency of hesitation pauses occurring between repeated words or phrases, when a false start was made, after connecting words, between the subject and a verb, and between a verb and its object. Hesitation pauses, if they are long and frequent, at these locations are, therefore, possible indicators of "non-native pauses".

CHAPTER V

IMPLICATIONS AND UNSOLVED PROBLEMS

What are the implications of these findings for the evaluation of fluency? What are the limitations of these findings? What problems are left unsolved? Each of these questions will be dicussed below.

The first finding of Experiment 1 partially verifies our hypothesis that fluency relates primarily to the temporal and sequential aspects of speech output. The verification is only partial because the variables we dealt with were quite limited in number. We proposed four factors that should be considered in the evaluation of fluency: 1) Speed of response, 2) Length and amount of hesitation pauses, 3) Number of repetitions and false starts, and 4) Logical coherence and the use of connectives. These factors supposedly constitute what is involved in the concept of fluency. The variables which were dealt with in Experiment 1 were only four (WPM, Percentage of Hesitation Pauses, Mean Hesitation Length, and Number of Sentences per Minute). These are related only to the factors 1) and 2). No variables belonging to the factors 3)and 4) were tested in relationship with the native speakers' impressions of fluency. In order to verify our hypothesis fully

many more variables should be analysed. The necessity to conduct such an extensive experiment to verify our hypothsis is obvious judging from the non-perfect correlations between the native subjects' impression of fluency and the four variables.

Another limitation of the first finding in Experiment 1 is related to the criterion measure of fluency used on the native speaker's impression of fluency. In other words, we do not know for sure whether the native subjects' impression of fluency is accurate. According to the reliability estimate (.95), the fifteen native subjects, as a group, were highly reliable. However, although they were instructed to rate "fluency," we cannot tell to which aspect of the recorded speech they really paid attention. Therefore, even if native speakers had a clear and consistent understanding of what it is to be fluent, it might be possible that their impressions of fluency are inadvertently influenced by other aspects of the non-native subjects' recorded speech. This kind of problem can be found in any study in which a concurrent validation of a test is done by correlating that test with another test which is "supposed" to be valid.

Even with these limitations, it can be said that our hypothesis about the nature of fluency was partly verified in Experiment 1, and that WPM, and Percentage of Hesitation Pauses can serve as reasonably valid measures of fluency.

Concerning the second finding in Experiment 1, again interpretation needs to be done with caution. The data indicates that fluency is more integrative in nature than any of the Grammar, Vocabulary, and Pronunciation scales. However, the high correlation of the Fluency scale with the other scales might be merely because it was difficult for the native subjects to distinguish fluency from the other three aspects. Again, the result should not be seen as decisive. Nevertheless, if our hypothesis is supported by the finding, it follows that fluency plays an important part in the overall evaluation of speaking. By focusing on fluency we would hope to measure a substantial part of the ability to speak.

Expeiment 2 was devoted to answering Question 3). The method used here to clarify the characteristics of "non-native pauses" was to compare the native and the non-native subjects' hesitation pauses. The different identified aspects of hesitation pauses were expected to be possible indicators of "non-native pauses". It should be recognized that this method allows us to infer only indirectly what are the characteritics of non-native pauses.

As might be expected, the non-native subject's hesitation pauses turned out to differ significantly from the native subjects both in overall amount and in frequency. However, these differences can give us only a gereral idea about the nature of non-native pauses. A more elaborate analysis

was carried out in relation to the locations of hesitation pauses. Hesitation pauses at the locations where there was a great difference between the two groups in the length of hesitation pauses could be used as useful indices of fluency.

This study attempted to solve only a very small portion of the problems concerning the evaluation of fluency. Many complicated problems concerning the exact factors which influence the native speaker's impressions of fluency and the relative importance of each factor are largely left unsolved.

First, although we suggested in Chapter II, 3.2, that four factors can cover a substantial portion of the concept of fluency, we are stll not sure whether other aspects of speech output influence the native speaker's impression of fluency. For example, in conversational situations, the use of phrases like ". . . something like that" might, if they are not excessive, make the speaker sound more fluent. Or, the use of rather long, complicated sentences and difficult words might make him sound more fluent in a lecture or a speech. Such factors, which might contribute to a native speaker's impressions of fluency of foreign language learners are not completely covered by the four factors we proposed. Thus, we regard our proposition as a tentative one and must seek to refine it.

Secondly, the weighting, i.e. the relative importance,

of the factors influencing the native speaker's impression is difficult to determine. In the present study we did find a high correlation of WPM and Percentage of Hesitation Pauses with the native subjects' impressionistic evaluation of fluency. As far as the subjects, the native judges, and the speech task required in the present study are concerned, we could say fairly confidently that the above two variables should be heavily weighted if they are to be used as indices of fluency. However, if, for example, the subjects are more proficient learners of English than those in the present study, we cannot conclude that these two factors would again be the most important. We cannot but hesitate to say any conclusive remarks about this. If such subjects did the same kind of a picture description and if their WPM ranged, say, from 100 to 150, we would expect that the weighting of each factor would change. Factors which were quite important for the native speakers to discriminate among learners whose WPM range from 50 to 100 might not be equally important when they discriminate learners whose WPM range from 100 to 150. We would suggest that the weightings of the factors must be varied depending upon whether we wish to evaluate the fluency of advanced learners or intermediate learners.

We would further suggest that situational features exert an even greater influence upon the relative importance

of each of the factors involved in the concept of fluency. The use of logical connectives (e.g. "therefore"), for instance, might be weighted less in an informal conversation than in a formal lecture. We would naturally have to change the weighting of WPM when we evaluate one's fluency depending upon whether he is newscasting or reciting a poem by Wordsworth.

Therefore, it would be impossible for us to have a single "fixed" set of weightings of factors influencing the notion of fluency, even though we succeeded to list up all those factors. If there were such a fixed set for evaluating fluency in a variety of situations, it would necessarily be only a general compromised set. What is crucial in the evaluation of fluency then would be to limit as much as possible the application of a particular criteria of fluency, i.e. a set of factors with their weighting, to a given group of testees, a given situational setting and register. This approach might be called "Evaluation for Specific Purposes". and would seem to be a comfortable excuse for not being able to devise any "general" measurement instrument of fluency. The contention here, however, is that we should try to develop, fist, reliable and valid criteria for fluency in specific contexts, and then hope to determine a general criteria as a common denominator.

Thirdly, we left largely untouched the problem of the

level of achievement. For example, how fast should one speak, or how little should one hesitate in order to be qualified as a fluent speaker of English as a foreign language? Ando(1979: 32) suggests 100 WPM as the acceptable level of fluency. This could be regarded as the level above which a foreign language learner sounds "fluent," but it is only a speculation from a subjective impression. Here again an acceptable level of fluency cannot be determined without reference to situational contexts.

Fourth, Sajavaara and Lehtonen (1978: 30) pose a question "Is a speaker who has very little to say but who is able to speak for long periods to be considered a fluent speaker or not?" This problem was given brief consideration in Chapter II, 3.2, but we did not make clear how rate of speech and information conveyed in a cerain time interact and cooperatively influence the native speaker's impression of fluency. In consideration of this problem the highest or lowest degree of redundancy permitted in spoken English has to be taken into account. If we neglect either of these, some problems might arise. For example, If we used only WPM as a criterion of fluency, a person who is quite good at filling the vacuum with appropriate words or phrases, yet who conveys little information could be judged more fluent than a person who can convey much greater amount of information in a shorter time but with long pauses. Similarly, if we used only

information conveyed in a certain time as a criterion, the former's skillful maneuvering in dealing with his temporal lack of information to convey, which does seem to influence the native speaker's impression of fluency,would be neglected. More research is needed concerning the influence of rate of speech and information conveyed within a certain time upon the impression of fluency and interaction of these two factors.

The last point to be made is about the proficiency level at which fluency is the most discriminating factor. In fact, it is premature to discuss this problem since the nature of fluency is not yet exactly understood. However, if we define fluency simply as WPM (though obviously too simplistic), we could speculate that fluency is probably not an important factor at a highly advanced level. The linguistic quality of speech output such as choice of vocabulary, would seem to be the most important factor in discriminating among students at such a level. A support for this view could be drawn from Adams (1978: 4). His study is concerned with the five factors (Accent, Comprehension, Fluency, Grammar, and Vocabulary) considered in the FSI Oral Interview test. His data showed that Fluency was not a discriminating factor at the highest proficiency level, but at intermediate or elementary levels. Not much is known about this issue and more research is needed.

CHAPTER VI

CONCLUSION

The main aspect of speaking ability usually tested seems to be "grammatical correctness" of speech output. The "facility" with which one speaks a foreign language is often neglected. If we consider fluency to be essential in the communicative situation, it is obvious that we cannot neglect this aspect when we evaluate speaking ability.

There are two ways to evaluate fluency in English as a foreign language. One is to use native speakers' intuition as the criterion. The other is to use an objective measures of quantifiable aspects of speech output with a consideration of the communicative situation and try to obtain a result identical with the result of the first method. The former method is a more valid way of evaluating fluency in that fluency means "that the communication processes meet the listener's expectation"(Sajavaara and Lehtonen 1978: 52), that the final criteria for fluency depend upon what the native speaker regards it to be. However, this method is subject to various limitations. Firstly, native speakers of English are not always available. Secondly, this kind of method does not fit in the teaching and learning context,

since what is evaluated by the native speaker is not explicitly known. Thirdly, unless we use a substantial number of native speakers as judges, the judgment on fluency might be biased by idiosyncracies.

It would be much better if we could possibly devise a measurement tool of fluency which is more reliable than a single native speaker's impression of fluency and more closely approximate the general consensus of the majority of native speakers about fluency. Such a measurement tool will enable language teachers and testers, both native and non-native, to evaluate learners' fluency with validity and reliability. In an attempt to achieve this ideal, the present study gives a very little more than promising hints.

Lehtonen (1981: 12) points out the following:

> To learn to speak fluently is above all to learn that, in different situations, different communicative registers are used. Fluency does not always imply an uninterupted flow of speech which is grammatically irreproachable. To be fluent in the right way, one has to know how to hesitate, how to be silent, how to self-correct, how to interrupt and how to complete one's expression. According to this definition of fluency, one must speak in a way that is expected by the linguistic community and that represents normal, acceptable and relaxed linguistic behaviour. Testing of this quality of speech is not possible by means of any instrumental method.

What he says here is quite right and indicates that we need to consider all of the intricacies of speech output in the evaluation of fluency. This is the case especially with the the evaluation of fluency of advanced learners of English,

whose levels of fluency cannot adequately be discriminated solely by such simple measures as WPM or amount of hesitation pauses.

The evaluation of fluency is far from a simple matter. However, does this imply that we should give up attempts to evaluate fluency, as Lehtonen claims? Definitely, not. Difficulty can never be foreign language teachers' or testers' excuse for completely abandoning the evaluation of fluency. It is possible, first, develop a measurement instrument, though it might be limited and insufficient, then to refine it. We need not be perfect from the beginning, though constant attention to the ideal should never be forgotten. Our belief is that the evaluation of fluency in English as a foreign language is difficult but possible.

Notes

1. This was quoted in Roy (1969: 21).

2. I owe my recognition and clarification of this point to the discussion about fluency held in Prfessor Matsumura's class (January 1982).

3. Actually, this compartmentalization of language skills is a great issue in the field of language testing and its validity needs close re-examination. However, in the present author's view, such compartmentalization is inevitable and even necessary for the sake of reliability in testing language skills. This justification, however, requires further discussion and verification.

   This is commonly referred to as the Discrete Point versus Integrative controversy and needs independent treatment such as could only be given in another thesis. About this controversy, see Oller (1979: 423-458, 1973: 184-199) or Ohtomo (1979: 9-11).

4. Quoted in Ingram (1977: 25). The following is the guide offered by Connolly and Sluckin (1971: 154):

   0.90-1.00 very high correlation; very strong relationship,

   0.70-0.90 high correlation; marked relationship,

   0.40-0.70 moderate correlation; substantial relationship,

   0.20-0.40 low correlation; a definite relationship but a small one,

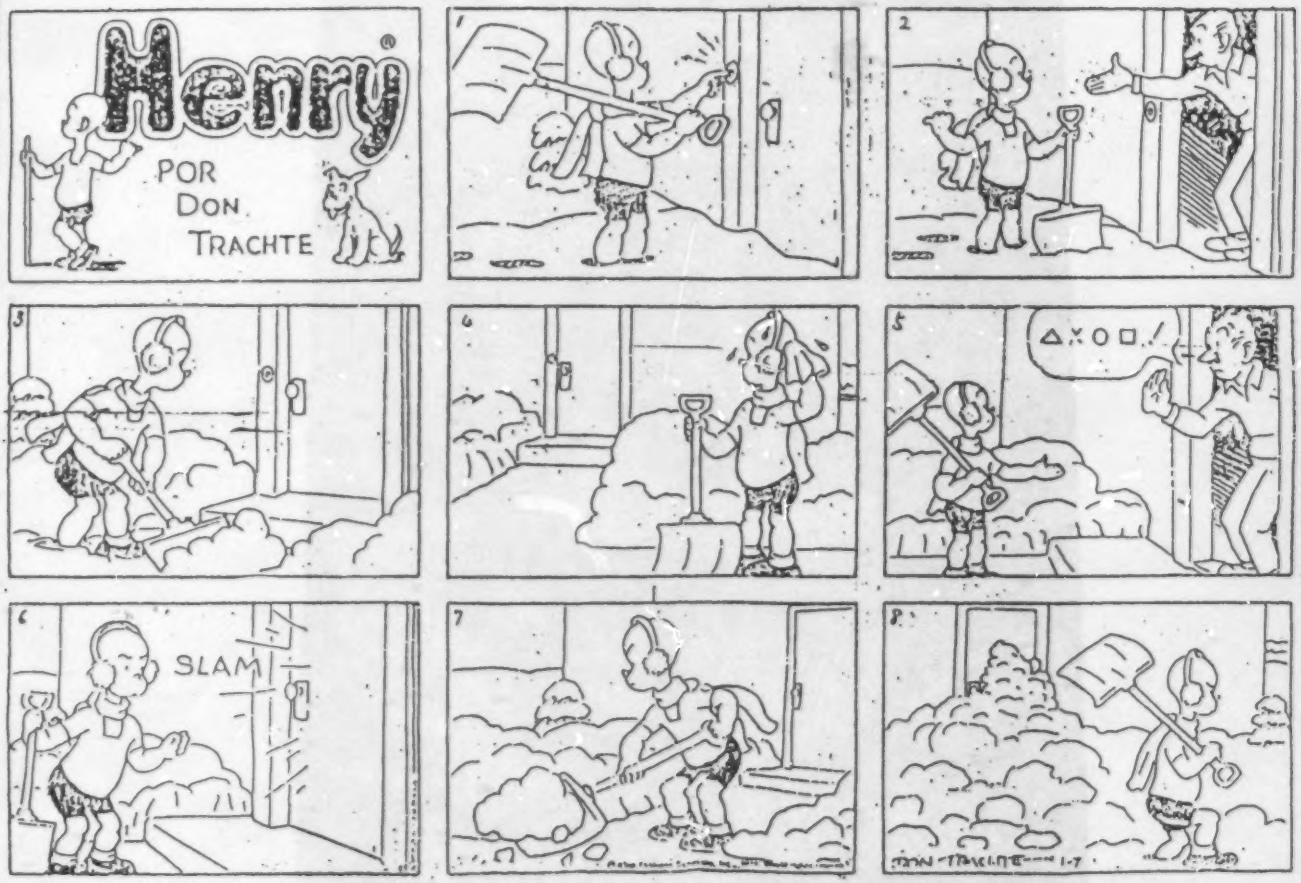
   0.20 and less slight correlation; relationship so small as to be negligible.

   They state that this is "a rough but useful guide to the degree of relationship indicated by the size of the coefficients".

5. Personal communication at the University of New Mexico, 1979.

APPENDIX I

A COMIC STRIP USED IN THE EXPERIMENTS



(This series of pictures came from the cartoon section of a Puerto Rican newspaper: El Mundo, Muñequito section, Feb.5, 1979)

BIBLIOGRAPHY

(A)

Adams, M.L. 1980. "Five Cooccurring Factors in Speaking Proficiency," Frith, J.R. (ed.) 1-6.

Allen, J.P.B. and A.Davies (eds.). 1977. The Edinburgh Course in Applied Linguistics: Testing and Experimental Methods. London: Oxford University Press.

Beardsmore, H.B. 1974. "Testing Oral Fluency," IRAL, 12, 4, 317-326.

_____ and A.Renkin. 1971. "A Test of Spoken English," IRAL, 9, 1, 1-11.

Blubaugh, J.A. 1969, "Effects of Positive and Negative Audience Feedback on Selected Variables of Speech Behavior," Speech Monographs, 36, 131-137.

Boomer, D.S. and A.T.Dittmann. 1962. "Hesitation Pauses and Juncture Pauses in Speech," Language and Speech, 5, 215-220.

Brière, E.J. 1971. "Are We Really Measuring Proficiency with Our Foreign Language Tests?" FLA, 4, 4, 385-391.

Brooks, N. 1964. Language and Language Learning: Theory and Practice, second edition. New York: Harcourt, Brace & World, Inc.

Brown, K. 1968. "Intelligibility," Davies, A. (ed.) 180-191.

Callaway, D.R. 1980. "Accent and the Evaluation of ESL Oral Proficiency," Oller, J.W. and K.Perkins (eds.) 102-115.

Carroll, B.J. 1980. Testing Communicative Performance: An Interim Study. Oxford: Pergamon Press.

Carroll, J.B. 1953. The Study of Language. Cambridge: Harvard University Press.

_____ 1964. Language and Thought. New Jersey: Prentice-Hall.

_____ 1968. "The Psychology of Language Testing," Davies, A. (ed.) 46-69.

_____ 1979. "Language Proficiency Tests Developed to the IEA International Study of Achievement in French as a Foreign Language," Spolsky, (ed.)

Catford, J.C. 1950. "Intelligibility," ELT, 5, 1, 7-15.

Chomsky, N. 1965. Aspects of the Theory of Syntax. Cambridge, Mass.: The M.I.T. Press.

Clark, J.L.D. 1972. Foreign Language Testing: Theory and Practice. Philadelphia: The Center for Curriculum Development.

_____ 1975. "Theoretical and Technical Considerations in Oral Proficiency Testing," Jones, R.L. and B.Spolsky (eds.) 10-28.

_____ (ed.) 1978. Direct Testing of Speaking Proficiency: Theory and Application. Princeton., New Jersey: Educational Testing Service.

_____ and S.S. Swinton. 1979. TOEFL Research Reports 4. An Exploration of Speaking Proficiency Measures in the TOEFL Context. Princeton, New Jersey: Educational Testing Service.

Clifford, R.T. 1980. "Foreign Service Institute Factor Scores and Global Rating," Frith (ed.) 27-30.

Cohen, A.D. 1980. Testing Language Ability in the Classroom. Rowley, Mass: Newbury House.

Connolly, T.G. and W.Sluckin. 1971. An Introduction to Statistics for the Social Sciences. third edition. London: Macmillan.

Crystal, D. and D.Davy. 1975. Advanced Conversational English. London: Longman.

Dalton, P. and W.J.Hardcastle. 1977. Disorders of Fluency and Their Effects on Communication. London: Edward Arnold.

Davies, A. 1968. "Introduction," Davies, A. (ed.) 1-18.

_____ (ed.) 1968. Language Testing Symposium: A Psycholinguistic Approach. London: Oxford University Press.

_____ 1978. "Language Testing - Part II," Language and Linguistics Abstracts, 11, 4, 215-231.

_____ 1978. "Language Testing (a survey article)," LTLA, 11, 3, 145-159.

Frith, J.R. (ed.) 1980. Measuring Spoken Language Proficiency. Washington, D.C.: Georgetown University Press.

Goldman Eisler, F. 1968. Psycholinguistics: Experiments in Spontaneous Speech. London: Academic Press.

Gove, P.B. (ed.) 1966. Webster's Third New International Dictionary of the English Language, Unabridged. Springfield, Mass.: G. & C. Merriam.

Grieve, D.W. 1968. "Tests in Education," Davies, A. (ed.) 133-150.

Harriss, D.P. 1969. Testing English as a Second Language. New York: McGraw-Hill.

Hendricks, D., C.Scholz, R.Spurling, M.Johnson and L. Vandenburg. 1980. "Oral Prodiciency Testing in an Intensive English Language Program," Oller, J.W. and K.Perkins (eds.) 77-90.

Henderson, A., F.Goldman Eisler and A.Skarbek. 1966. "Sequential Temporal Patterns in Spontaneous Speech," Language and Speech, 9, 4, 207-216.

Hieke, A.E. 1981. "A Content-Processing View of Hesitation Phenomenon," Language and Speech, 24, 2, 147-160.

Ingram, E. 1968. "Attainment and Diagnostic Testing," Davies, A. (ed.) 70-97.

_____ 1977. "Basic Concepts in Testing," Allen, J.P.B. and A.Davies (eds.) 11-37.

Jones, R.L. and B. Spolsky (eds.) 1975. Testing Language Proficiency. Arlington, Virginia: Center for Applied Linguistics.

Katz, J.J. and J.A.Fodor. 1964. "The Structure of a Semantic Theory," Katz, J.J. and J.A.Fodor (eds.) The Structure of Language. Englewood Cliffs: Prentice Hall, 481-482.

Lado, R. 1961. Language Testing: The Construction and Use of Foreign Language Tests. London: Longman.

Leeson, R. 1975. Fluency and Language Teaching. London: Longman.

Lehtonen, J. 1978. "Problem of Measuring Fluency and Normal Rate of Speech," Jyäskylä, Finland: Department of Phonetics and Linguistics, University of Jyväskylä. Quoted in Cohen, A.D. 1980. 122.

Lehtonen, J. 1981. "Problems of Measuring Fluency and Normal Rate of Speech," Savard, J.G. and L.Laforge (eds.) Proceedings of the 5th Congress of AILA, Montreal, August 1978, 322-332.

Lowe, P.Jr. and R.T.Clifford. 1980. "Developing an Indirect Measure of Overall Oral Proficiency," Paper presented at the TESOL Convention, San Francisco, California. March, 1980.

Madsen, H.S. and R.L.Jones. 1980. "A Survey of Oral Proficiency Tests: Phase II," A Paper read at The Colloquium on Validation of Oral Proficiency Tests, March, San Francisco, California.

Maclay, H. and C.E. Osgood. 1959. "Hesitation Phenomena in Spontaneous English Speech," Word, 15, 1, 19-44.

Morrow, K. 1979. "Communicative Language Testing: Revolution or Evolution?" Brumfit, C,J. and K.Johnson (eds.) The Communicative Approach to Language Teaching. London: Oxford University Press.

Mullen, Karen A. 1978. "Direct Evaluation of Second-language Proficiency: The Effect of Rater and Scale in Oral Interviews," LL, 28, 2, 301-308.

Naiman, N., Frohlick M., Stern H.H. and Todesco A. 1978. The Good Language Learner. The Ontario Institute for Studies in Education.

Oller, J.W. 1979. Language Tests at School. London: Longman.

_____ and F.B.Hinofotis. 1980. "Two Mutually Exclusive Hypotheses about Second Language Ability: Indivisible or Partially Divisible Competence," Oller, J.W. and K.Perkins (eds.) 13-23.

_____ and K.Perkins (eds.) 1980. Research in Language Testing. Rowley, Mass.: Newbury House.

Osgood, C.E. and T.A.Sebeok (eds.) 1965. Psycholinguistics: A Surbey of Theory and Reseach Problems. Bloomington: Indiana University Press.

Ozasa, T. 1979. "Characteristic Features of Spoken English as Spoken by Japanese and Native Speakers," Kagoshima University Eigo Eibungaku Ronshu. 10, 187-229.

Perren, C.E. 1968. "Testing Spoken Language: Some Unsolved Problems," Davies, A. (ed.) 107-116.

Rivers, W.M. 1968. Teaching Foreign-Language Skills. Chicago: The University of Chicago Press.

Robinson, P. 1971. "Oral Expression Tests: 1," ELT, 25, 2, 151-155.

_____ 1971. "Oral Expression Tests: 2," ELT, 25, 3, 260-266.

Roy, R. 1969. "Rate of Output - - - Factor of Oral Proficiency," CMLR, 26, 1, 14-22.

Sajavaara, K. and J.Lehtonen. 1978. "Spoken Language and the Concept of Fluency," Lautamatti, L. and P.Lindqvist (eds.) Focus on Spoken Language. Language Centre News, Special Issue 2. University of Jyväskylä. 23-57.

Seward, B.H. 1973. "Measuring Oral Production in EFL," ELT, 28, 1, 77-80.

Shannon, C.E. 1951. "Prediction and Entropy of Printed English," The Bell System Technical Journal, 30.

Spolsky, B. (ed.) 1979. Papers in Applied Linguistics: Advances in Language Testing Series: 1, Some Major Tests, Arlington, Virginia: Center for Applied Linguistics.

Stack, E.M. 1966. The Language Laboratory and Modern Language Teaching. Revised edition. London: Oxford University Press.

Stevick, E.W. 1976. Memory, Meaning and Method: Some Psychological Perspectives on Language Learning. Rowley, Mass.: Newbury House.

Sykes, J.B. 1976. The Concise Oxford Dictionary of Current English. Sixth edition. London: Oxford University Press.

Takenaka, T. 1979. "Some Considerations on the Fluency in a Foreign Language: From a Communicative Point of View," The Journal of Hiroshima University Graduate School. 5, 110-116.

Townson, M. 1973. "Testing Oral Skills at University Level," ELT, 27, 2, 199-205.

Valette, R.M. 1977. Modern Language Testing. Second edition. New York: Harcourt Brace Jovanovich.

Welkowitz, J., R.B.Ewen and J.Cohen. 1976. Introductory Statistics for the Behabioral Sciences. second edition. New York: Academic Press.

Yorozuya, R. and J.W.Oller,Jr. 1980. "Oral Proficiency Scales: Construct Validity and the Halo Effect," LL, 30, 1, 135-153.

安藤昭一，比嘉正範，西村嘉太郎．　1979．「Japanese English は目標か結果か」『英語教育』，28，1，30-35．

大友賢二．　1981．「英語能力の構造」『英語教育』，30，9．9-11

貝瀬千章．　1980．「速度ではかる英語力」，The English Journal，10，11，16-18．

沢登春仁．1980．「なぜ日本人は話しべたか」，『英語教育』28，13，9-11．

羽鳥博愛，松畑熙一．　1980．『学習者中心の英語教育』東京：大修館．

松畑熙一．　1974．「評価活動と学習指導 —— speaking を中心として ——」『九州英語教育学会紀要』．3，7-17．

土屋澄男．　1979．「英語科におけるテストと評価 (7)：話すことのテスト」．『現代英語教育』．14，7，12-13．

柳井智彦．　1979．「大学生のスピーキング運用力：録音データの分析から」『教育学研究紀要』25，237-8．